

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: December 17, 2004, 00:05:37 ; Search time 3793 Seconds

(without alignments)  
2231.705 Million cell updates/sec

Title: US-09-701-674a-23

Perfect score: 179  
Sequence: 1 MGTGSLSDVEDLQEVEMLESC.....PSSDLKEVTVASRLCGTTAS 179

Scoring table:

OLIGO  
Xgapop 60.0 , Xgapext 60.0  
Ygapop 60.0 , Ygapext 60.0  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 4526729 seqs, 23644849745 residues

Word size: 1

Total number of hits satisfying chosen parameters: 9045947

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame\_plus\_p2n.model -DRV=xlh  
-Q=/cgn2\_1/USPTO.spool\_h/US09701674/runat\_15122004.100530.28306/app.query.fasta\_1.327  
-DB=GenEmbl -OPMT=fastap -SUFFIX=oligo.rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi -LIST=45  
-DOCALLCN=200 -THR SCORE=quality -THR MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc  
-NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09701674.CGN 1.1.3731 @runat\_15122004.100530.28306 -NCEU=6 -ICPU=3  
-NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -USPSLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

GenEmbl.\*  
1: gb\_ba.\*  
2: gb\_hgt.\*  
3: gb\_in.\*  
4: gb\_cm.\*  
5: gb\_cv.\*  
6: gb\_pat.\*  
7: gb\_ph.\*  
8: gb\_pl.\*  
9: gb\_pr.\*  
10: gb\_ro.\*  
11: gb\_sts.\*  
12: gb\_sy.\*  
13: gb\_un.\*  
14: gb\_vi.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result , No.	Score	Query Match	Length	DB ID	Description
1	179	100.0	537	9	CR450293 Homo sapi
2	179	100.0	1272	6	BD211771 Proteins
3	179	100.0	3231	9	BC025697 Homo sapi
4	166	92.7	1259	6	CQ718038 Sequence

5	150	83.8	25186	9	AL356109	Human DNA
6	120	67.0	1202	10	AF047418	Mus muscu
7	120	67.0	1217	10	AF029753	Mus muscu
8	120	67.0	1220	10	AB009453	Mus muscu
9	120	67.0	1240	10	AF035717	Mus muscu
10	120	67.0	1245	10	BC053525	Mus muscu
11	120	67.0	1267	10	AF036945	Mus muscu
12	118	65.9	1254	9	AF035718	Homo sapi
13	109	60.9	187332	2	AC101292	Mus muscu
14	109	60.9	241544	2	AC115183	Rattus no
15	98	54.7	1164	5	BC932800	Gallus ga
16	96	53.6	1257	9	AF047419	Homo sapi
17	88	49.2	965	5	BC073597	Xenopus l
18	88	49.2	1059	5	AY660871	Xenopus l
19	86	48.0	513	10	AF061752	Rattus no
20	77	43.0	124070	5	BC530074	zebrafish
21	77	43.0	231207	2	CR293516	Danio rer
22	67	37.4	189465	2	CR318665	Danio rer
23	61	34.1	697	6	AX333347	Sequence
24	61	34.1	697	6	AX333763	Sequence
25	36	20.1	240	6	AR270465	Sequence
26	28	15.6	606	10	AF108216	Mus muscu
27	28	15.6	621	9	BT007169	Homo sapi
28	28	15.6	621	12	BT008287	Synthetic
29	28	15.6	1716	9	AF087036	Homo sapi
30	28	15.6	1718	10	AF087035	Mus muscu
31	28	15.6	1893	9	AF060154	Homo sapi
32	28	15.6	1917	6	CQ717187	Sequence
33	28	15.6	2031	9	BC006313	Homo sapi
34	28	15.6	2040	9	BC067827	Homo sapi
35	28	15.6	2382	6	AX281674	Sequence
36	28	15.6	163698	10	AC115876	Mus muscu
37	28	15.6	170668	9	AC104012	Homo sapi
38	28	15.6	18157	10	AC121789	Mus muscu
39	28	15.6	196331	9	AC028867	Homo sapi
40	28	15.6	225399	2	AC113255	Rattus no
41	25	14.0	222122	2	CR450806	Danio rer
42	21	11.7	107	11	AB094420	Sus scro
43	19	10.6	144007	2	AC141743	Apis mell
44	18	10.1	65	6	CQ530499	Sequence
45	16	8.9	101	11	BV184299	sqnm14420

## ALIGNMENTS

CR450293 537 bp mRNA linear PRI 18-MAY-2004  
Homo sapiens full open reading frame cDNA clone RZPD0834B11D for  
gene TCF21, transcription factor 21; complete cds; without  
stopcodon.

CR450293  
CR450293.1 GI:47496532  
Full ORF shuttle clone, Gateway(TM), complete cds.  
Homo sapiens (human)

CR450293  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 537)  
Ebert.L., Schick.M., Neubert.P., Schattten.R., Henze.S. and Korn.B.  
Cloning of human full open reading frames in Gateway(TM) system  
entry vector (pDONR201)

Unpublished

2 (bases 1 to 537)

Ebert.L., Schick.M.,

Direct Submission

Submitted (18-MAY-2004) RZPD Deutsches Ressourcenzentrum fuer

Genomforschung GmbH, Im Neuenheimer Feld 580, D-69120 Heidelberg,

Germany

RZPD; RZPD0834B11D, ORFNO 76

www.rzpd.de/cgi-bin/products/cl.cgi?cloneID=RZPD0834B11D RZPDLIB;

Human Full ORF Clones Gateway(TM) - RZPD (kan-resist.) RZPD LIB No.

834

www.rzpd.de/cgi-bin/products/showlib.pl.cgi/response?libNo=834  
www.rzpd.de/products/orfclones/

Contact: Ina Rolfes  
RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH,  
Heubnerweg 6, D-14059 Berlin, Germany  
Tel: +49 30 32639 101  
Fax: +49 30 32639 111  
www.rzpd.de

This clone is available from RZPD;  
contact RZPD (customer.service@rzpd.de) for further information.  
This CDS clone is a part of a collection of human full length  
expression clones generated by RZPD.

This CDS has been cloned without stopcodon.  
The CDS has been inserted into pDONR201 via a BP Clonase(TM)  
reaction. Additional sequence has been added in front of the start  
codon (ATG): att. AAAAAA GGT GGC ACC CCT GGT CCA GGT (ATG)  
After the last codon additional sequence has been added: CCA GGC  
CCA GGC GGC G in front of the 3' att site (AC CCA GCT TTC TT).  
Compared to the reference sequence NM\_003206 we did not find any  
amino acid exchanges.

Clone distribution: <http://www.rzpd.de/products/orfclones/>.

#### FEATURES

source  
1..537  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="RZPD08348111D"  
/clone\_lib="Human Full ORF Clones Gateway(TM) - RZPD"  
/lab\_host="DH10B"  
/note="Vector: pDONR201, Site\_1: attP1; Site\_2: attP2"  
gene  
1..537  
/genes="TCF21"  
CDS  
1..537  
/genes="TCF21"  
/codon\_start=1  
/protein\_id="CAG29289.1"  
/db\_xref="GI:47496533"  
/translation="MSGSLSDVEDLQEVMECDGLKWDNSKEFVTSNESTERSSNC  
ENSPQKGGGKRRKATPKKSPGVSQEGKQVORNAANARERARVLSKAPSLR  
KTLPWPPPTKLKSLDITLRLASSYIAHLRLANDKYENYIHPVNLTPWTFVWAGP  
ESDLKEVVTASRLCGTAS"

#### ORIGIN

Alignment Scores:  
Pred. No.: 2,71e-184 Length: 537  
Score: 179.00 Matches: 179  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 9 Gaps: 0

US-09-701-674A-23 (1-179) x CR450293 (1-537)

Qy 1 MetSerThrGlySerLeuSerAspValGluAspLeuGlnGluValGluMetLeuGluCys 20  
Db 1 ATGTCACCGGCTCCCTCAGCATGTGGAGGACCTTCAAGAGGTGAGATGTGGAATGT 60  
Qy 21 AspGlyLeuLysMetAspSerAsnLysGluPheValThrSerAsnGluSerThrGluGlu 40  
Db 61 GACGGGTGTGAATGTGAACTCGAACAGGAATTTGTGACTTCCACGAGACCGAGGAG 120  
Qy 41 SerSerAsnGlyGlnGlnGlySerProGlnLysGlyArgGlyGlyLeuGlyLysArg 60  
Db 121 AGCTCAACTCGGAGAAATGGTCTCCAGAGGGCGCGGCGCTGGCGCAAGAGGAGG 180  
Qy 61 LysAlaProThrLysLysSerProLeuSerGlyValSerGlnGluGlyLysValGln 80  
Db 181 AAGGCCGCCCAAGAGAGAGCCCTTGGAGCGGGGTGAGCAGAGGGGAGCAGGTCCAG 240  
Qy 81 ArgAsnAlaAlaAsnAlaArgGluArgAlaArgMetArgValLeuSerLysAlaPheSer 100  
Db 241 CGCAACGGCGCCACACGGCGGAGAGCGGGCCCGCATCGAGAGTCTGAGCAAGCCCTTCC 300

Qy 101 ArgGlyLysThrThrLeuProTrrValProAspThrLysLysLeuSerLysLeuAspThr 120  
Db 301 AGATCAGACACCCCTGCGCTGGTGGTCCCGCCGACACCAAGCTCTCCAGCTGGACAG 360  
Qy 121 LeuArgLeuAlaSerSerTyrlleAlaHisLeuArgGlnlleleuAlaAsnAspLysTyr 140  
Db 361 CTCAGGCTGGCGTCCAGCTACATCGCCCACTTGAGGCAGATCCTGGCTAACGACAAATAC 420  
Qy 141 GluAsnGlyTyrlleHisProValAsnLeuThrTrrProPheMetValAlaGlyLysPro 160  
Db 421 GAGAACGGGTACATTCACCGCGTCACTGACGTGGCCCTTTATGTGGCCGCGGAACCC 480  
Qy 161 GluSerAspLeuLysGluValValThrAlaSerArgLeuCysGlyThrThrAlaSer 179  
Db 481 GAGAGTGACCTGAAGAAGTGGTGACCGCGAGCGCTTATGTGAACCAACCGCGGTCC 537  
RESULT 2  
BD211771 1272 bp DNA linear PAT 17-JUL-2003  
LOCUS Proteins regulating gene expression.  
DEFINITION BD211771  
ACCESSION BD211771  
VERSION 1 GI:33021541  
KEYWORDS JP 2002517246-A/23  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 1272)  
AUTHORS Lal, P., Yue, H., Tang, Y.T., Hillman, J.L., Bandman, O., Corley, N.C.,  
Guegler, K.J., Gorgone, G.A., Baughn, M.R., Patterson, C. and Lu, D.A.M.  
TITLE Proteins regulating gene expression  
JOURNAL Patent: JP 2002517246-A 23 18-JUN-2002;  
COMMENT OS Homo sapiens (human)  
INCYTE PHARMACEUTICALS INC  
PN JP 2002517246-A/23  
PF 18-JUN-2002  
PD 11-JUN-1999 JP 2000553586  
PR 12-JUN-1998 US 60/089029, 29-JUL-1998 US 60/094575 PR  
PI PREETI LAL, HENRY YUE, Y TOM TANG, JENNIFER L HILLMAN, OIGA PI  
BANDMAN,  
PI NEIL C CORLEY, KARL J GUEGLER, GINA A GORGONE, MARIAH R BAUGHN,  
PI CHANDRA PATTERSON, DYUNG AINA M LU  
PC C12N15/00,A61K38/00,A61K39/395,A61K39/395,A61K45/00,A61P15/00,  
PC A61P25/00,  
PC A61P35/00,A61P37/02,A61P43/00,C07K14/47,C07K16/18,C12N15/00,  
PC A61K37/02  
CC Incyte Clone No.: 3340296  
FH Key Location/Qualifiers  
FT source 1..1272  
FT /organism="Homo sapiens (human)".  
FEATURES  
source  
1..1272  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"

#### ORIGIN

Alignment Scores:  
Pred. No.: 6,16e-184 Length: 1272  
Score: 179.00 Matches: 179  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 6 Gaps: 0

US-09-701-674A-23 (1-179) x BD211771 (1-1272)

Qy 1 MetSerThrGlySerLeuSerAspValGluAspLeuGlnGluValGluMetLeuGluCys 20  
Db 268 ATGTCCACCGGCTCCCTCAGCATGTGGAGGACCTTCAAGAGGTGAGATGTGGAATGT 327  
Qy 21 AspGlyLeuLysMetAspSerAsnLysGluPheValThrSerAsnGluSerThrGluGlu 40

Db 328 GACGGTTGAAATGATTTCGAAACAGGAATTTGTGACTTCCACGAGACCCAGGAG 387  
 Qy 41 SerSerAsnCysGluAsnGlySerProGlnLysGlyArgGlyGlyLeuGlyLysArgAG 60  
 Db 388 AGCTCCAACTGCGAGATGGTCTCCACAGAGGCGCGGGGCTCGGCAAGAGGAG 447  
 Qy 61 LysAlaProThrLysLysSerProLeuSerGlyValSerGlnGluGlyLysGlnValGln 80  
 Db 448 AAGCGCCCCACCAAGAGAGCCCTGAGCGGGGTGAGCCAGAGGAGGAGAGGTCCAG 507  
 Qy 81 ArgAsnAlaAlaAsnAlaArgGluArgAlaArgMetArgValLeuSerLysAlaPheSer 100  
 Db 508 CGCAGAGCCGCCAACCGGAGGCGGCGCGCATGCGAGTCTGAGCAAGCCCTTCTCC 567  
 Qy 101 ArgLeuLysThrThrLeuProThrValProProAspThrLysLeuSerLysLeuAspThr 120  
 Db 568 AGACTCAAGACCAACCTGCGGTGGTCCCGCCGACACCAAGCTCTCCAAGCTGGACAG 627  
 Qy 121 LeuArgLeuAlaSerSerTyrIleAlaHisLeuArgGlnIleLeuAlaAsnAspLysTyr 140  
 Db 628 CTGAGGTGGTCCAGCTACATCGCCCACTTGAAGCAGATCTGCTGAGCAAGCCCTTCTCC 687  
 Qy 141 GluAsnGlyTyrIleHisProValAsnLeuThrTrpProPheMetValAlaGlyLysPro 160  
 Db 688 GAGACGGGTACATTCACCGGTCAACCTGACGTGGCCCTTTATGGTGGCCGGAAACCC 747  
 Qy 161 GluSerAspLeuLysGluValThrAlaSerArgLeuCysGlyThrThrAlaSer 179  
 Db 748 GAGAGTGACCTGAAGAAGTGGTGACCGGAGCGGCTTATGTGGAACCAACCGCGTCC 804

## RESULT 3

BC025697

LOCUS

DEFINITION Homo sapiens transcription factor 21, transcript variant 1, mRNA  
 (CDNA clone MGC:34534 IMAGE:5221575), complete cds.

ACCESSION

BC025697

VERSION

MGC.

KEYWORDS

MGC.

SOURCE

Homo sapiens (human)

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 3231)

AUTHORS

Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D., Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, J., Max, S.I., Wang, J., Hsieh, F., Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Scheetz, T.E., Brownstein, M.J., Ussin, T.B., Casavant, T.L., Carninci, P., Prange, C., Raha, S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullen, S.J., Bosak, S.A., McEwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W., Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahey, J., Helton, E., Kettman, M., Madan, A., Rodriguez, S., Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Small, D.E., Schnerch, A., Schein, J.E., Jones, S.J., and Marra, M.A.

Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences

Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

12477932

2 (bases 1 to 3231)

AUTHORS

Strausberg, R.

Direct Submission

Submitted (06-MAR-2002) National Institutes of Health, Mammalian

Gene Collection (MGC), Cancer Genomics Office, National Cancer

Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,

USA

TITLE

JOURNAL

PUBMED

REFERENCE

2 (bases 1 to 3231)

AUTHORS

Strausberg, R.

Direct Submission

Submitted (06-MAR-2002) National Institutes of Health, Mammalian

Gene Collection (MGC), Cancer Genomics Office, National Cancer

Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,

USA

## REMARK

COMMENT

NIH-MGC Project URL: http://mgs.nci.nih.gov  
 Contact: MGC help desk  
 Email: cgabs-re@mail.nih.gov  
 Tissue Procurement: Life Technologies, Inc.  
 cDNA Library Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: National Institutes of Health Intramural  
 Sequencing Center (NISC),  
 Gaithersburg, Maryland;  
 Web site: http://www.nisc.nih.gov/  
 Contact: nisc\_mgc@nhgri.nih.gov  
 Akhter, N., Ayele, K., Beckstrom-Sternberg, S.M., Benjamin, B.,  
 Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S.,  
 Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghighi, P.,  
 Hansen, N., Ho, S.-L., Karlins, E., Kwong, P., Laric, P., Legaspi, R.,  
 Maduro, Q.L., Masiello, C., Maskeri, B., Mastrian, S.D., McCloskey, J.C.,  
 McDowell, J., Pearson, R., Stantripop, S., Thomas, P.J., Touchman, J.W.,  
 Tsurgoev, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L.,  
 Young, A., Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found  
 through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov  
 Series: IRAK Plate: 49 Row: k Column: 7  
 This clone was selected for full length sequencing because it  
 passed the following selection criteria: matched mRNA gi: 4507394.

## FEATURES

Location/Qualifiers

1..3231

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="MGC:34534 IMAGE:5221575"

/tissue types="Pancreas, Spleen, adult pooled"

/clone\_lib="NIH MGC\_120"

/lab\_host="DH10B"

/note="vector: pCMV-SPORT6"

1..3231

/gene="TCF21"

/note="synonym: POD1"

/db\_xref="LocusID:6943"

/db\_xref="MIM:603306"

260..799

/gene="TCF21"

/codon\_start=1

/product="transcription factor 21"

/protein\_id="AAH25697.1"

/db\_xref="GI:19344016"

/db\_xref="LocusID:6943"

/db\_xref="MIM:603306"

/translation="MSTGSLSDVEDLQEVEMLECDGLKMDSNKEFTVTSNEBSSNC  
 ENGSPQKRGGLGKRRKAPTKSPGVSQEGKQVQRNANARERARMVLKAFSL  
 KTLPLWPVPDTKLKSLDTLRLASSYIAHLRQILANDKYENGIHPVNLTPFWVAGKP  
 ESDLKEVVTASRLCGTAS"

ORIGIN

Alignment Scores:

Pred. No.:

Score:

Percent Similarity:

Best Local Similarity:

Query Match:

DB:

US-09-701-674A-23 (1-179) x BC025697 (1-3231)

Qy 1 MetSerThrGlySerLeuSerAspValGluAspLeuGlnGluValGluMetLeuGluCys 20

Db 260 ATGTCACCGGCTCCCTCAGCATGTGAGGACCTTCAGAGGTGGAGATGTTGAATGT 319

Qy 21 AspGlyLeuLysMetAspSerAsnLysGluPheValThrSerAsnGluSerThrGluGlu 40

Db 320 GACGGTTGAAATGAGTATCGAACAGGAATTTGTGACTTCCACGAGACCCAGGAG 379

Qy 41 SerSerAsnCysGluAsnGlySerProGlnLysGlyArgGlyGlyLeuGlyLysArg 60

```

Db      380 AGTCTCACTGCGAGAAATGGTCTCCCGAGAGGGCGCGGGCTGGCGAGAGGAGG 439
Qy      61  LyAlaProThrLysLysSerProLeuSerGlyValSerGlnGluGlyLysGlnValGln 80
Db      440 AAGCGCGCCACCAAGAAGAGCCCTCGAGCGGGGTGAGCCAGAGGGGAGGAGGTCCAG 499
Qy      81  ArgAsnAlaAlaAsnAlaArgGluArgAlaArgMetArgValLeuSerLysAlaPheSer 100
Db      500 CGCAAGCCGCCAACCGGAGAGCGGCCCGCATCGAGTGTCTGAGCAAGAGCCCTTCTCC 559
Qy      101 ArgLeuLysThrThrLeuProTrrValProProAspThrLysLeuSerLysLeuAspThr 120
Db      560 AGACTCAAGACCAACCTGCGCTGGGTGCGGCCCGCACCAAGCTCTCCAAAGTGGACACG 619
Qy      121 LeuArgLeuAlaSerSerTyrIleAlaHisLeuArgGlnIleLeuAlaAsnAspLysTyr 140
Db      620 CTGAGGTGGCGGTCAGCTACATCGCCCACTTGAGGCAGATCTTGCTCAACGACAAATAC 679
Qy      141 GluAsnGlyTyrIleHisProValAsnLeuThrTrpProPheMetValAlaGlyLysPro 160
Db      680 GAGAACGGGTACATTACCCGGTCAACCTGACGTGCGCCCTTTATGCTGCGCGGAAACCC 739
Qy      161 GluSerAspLeuLysGluValValThrAlaSerArgLeuCysGlyThrThrAlaSer 179
Db      740 GAGAGTGACCTGAAGAAGTGGTGACCGAGCGCGCTTATGTGGAACCAACCGCGTCC 796

RESULT 4
CQ718038      1259 bp      DNA      linear      PAT 03-FEB-2004
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1
Venter, C.J., Adams, M.C., Li, P.W. and Myers, E.W.
Kits, such as nucleic acid arrays, comprising a majority of
humanexons or transcripts, for detecting expression and other uses
thereof
JOURNAL
Patent: WO 02068579-A 3972 06-SEP-2002;
PE Corporation (NY) (US)
FEATURES
source
1. 1259
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
ORIGIN
Alignment Scores:
Pred. No.:      7,76e-170      Length:      1259
Score:      166.00      Matches:      179
Percent Similarity:      99.44%      Conservative:      0
Best Local Similarity:      99.44%      Mismatches:      0
Query Match:      92.74%      Indels:      1
DB:      6      Gaps:      0

US-09-701-674A-23 (1-179) x CQ718038 (1-1259)
Qy      1  MetSerThrGlySerLeuSerAspValGluAspLeuGlnGluValGluMetLeuGluCys 20
Db      264 ATGTCCACCGGCTCCCTCAGCATGTGGAGGACCTTCAGAGGGTGAGATGTGGAAATGT 323
Qy      21  AspGlyLeuLysMetAspSerAsnLysGluPheValThrSerAsnGluSerThrGluGlu 40
Db      324 GACGGGTGAAAATGGATTGCAACAAAGGAATTTGTGACTTCCAAACGAGAGCACCAGGAG 383
Qy      41  SerSerAsnCysGluAsnGlySerProGlnLysGlyArgGlyGlyLeuGlyLysArgArg 60
Db      384 AGCTCCAACTGCAGAAATGGTCTCCCGAGAGGGCGCGGGCTGGCGCAAGAGGAGG 443

```

```

Qy      61  LyAlaProThrLysLysSerProLeuSerGlyValSerGlnGluGlyLysGlnValGln 80
Db      444 AAGCGCGCCACCAAGAAGAGCCCTCGAGCGGGGTGAGCCAGAGGGGAGGAGGTCCAG 503
Qy      81  ArgAsnAlaAlaAsnAlaArgGluArgAlaArgMetArgValLeuSerLysAlaPheSe 100
Db      504 CGCAACGCCGCCAACCGGAGAGCGGGGCCCGCATCGAGTGTCTGAGCAAGAGCCCTTCTC 563
Qy      100 rArgLeuLysThrThrLeuProTrrValProProAspThrLysLeuSerLysLeuAspTh 120
Db      564 CAGACTCAAGACCAACCTGCGCTGGGTGCGGCCCGCACCAAGCTCTCCAAAGTGGACAC 623
Qy      120 rLeuArgLeuAlaSerSerTyrIleAlaHisLeuArgGlnIleLeuAlaAsnAspLys 140
Db      624 GCTCAGGCTGGCGTCCAGCTACATCGCCCACTTGAGGCAGATCTTGCTCAACGACAAAT 683
Qy      140 rGluAsnGlyTyrIleHisProValAsnLeuThrTrpProPheMetValAlaGlyLysPr 160
Db      684 CGAAGACGGGTACATTACCCGGTCAACCTGACGTGCGCCCTTTATGCTGCGCGGAAACC 743
Qy      160 oGluSerAspLeuLysGluValValThrAlaSerArgLeuCysGlyThrThrAlaSer 179
Db      744 CGAGAGTGACCTGAAGAAGTGGTGACCGGAGCGCGCTTATGTGGAACCAACCGCGTCC 801

AL356109      25186 bp      DNA      linear      PRI 23-JAN-2001
Human DNA sequence from clone RP11-373A10 on chromosome 6. Contains
ESTs, STSs, GSSs and CpG islands. Contains the TCF21 gene for
transcription factor 21, complete sequence.
AL356109      8 GI:10443434
HTG; CpG island; TCF21; transcription factor 21.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 25186)
Wallis, J.
Direct Submission
Submitted (15-DEC-2000) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Oct 1, 2000 this sequence version replaced gi:10186519.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C/elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping
Group. Further information can be found at
http://www.sanger.ac.uk/Hgp/Chr6
IMPORTANT: This sequence is not the entire insert of clone
RP11-373A10 It may be shorter because we sequence overlapping
sections only once, except for a 100 base overlap.
The true left end of clone RP1-73H22 is at 25087 in this sequence.
The true right end of clone RP4-662A9 is at 100 in this sequence.
This sequence has been finished according to sequence map criteria
as follows. An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated repeat sequence elements. Where the sequence is
ambiguous, there is an annotation using the 'unsure' feature key.
RP11-373A10 is from the library RPI1-11.2 constructed by the group
of Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm

```





repeat\_region /note="MLT1H repeat: matches 1. .286 of consensus"  
24559. 25036  
misc\_feature /note="MLT1D repeat: matches 1. .505 of consensus"  
complement(24819. .25186)  
/note="match: STS: Em:G30455"

## ORIGIN

Alignment Scores:  
Pred. No.: 3 08e-151 Length: 25186  
Score: 150.00 Matches: 150  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 83.80% Indels: 0  
DB: 9 Gaps: 0

US-09-701-674A-23 (1-179) x ALJ356109 (1-25186)

Qy 1 MetSerThrGlySerLeuSerAspValGluAspLeuGlnGluValGluMetLeuGluCys 20  
Db 13182 ATGTCACCGGCTCCCTCAGCGATGTGGAGACCTTCAAGAGGTGGAGATGTTGAATGT 13241  
Qy 21 AspGlyLeuLysMetAspSerAsnLysGluPheValThrSerAsnGluSerThrGluGlu 40  
Db 13242 GACGGTGTGAATGATTCGAACAGGAATTTGTGACTTCCACGAGAGCACCAGGAG 13301  
Qy 41 SerSerAsnCysGluAsnGlySerProGlnLysGlyArgGlyLeuGlyLysArgArg 60  
Db 13302 AGCTCCAACTCGAGAATGGTCTCCACAGAGGCGCGGGCGCTGGGCAAGAGGAGG 13361  
Qy 61 LysAlaProThrLysLysSerProLeuSerGlyValSerGlnGluLysGlnValGln 80  
Db 13362 AAGCGGCCCAACAGAGAGCCCTTGGAGGGGTGAGCAGAGGGAGGAGGTCACG 13421  
Qy 81 ArgAsnAlaAlaAsnAlaArgGluArgAlaArgMetArgValLeuSerLysAlaPheSer 100  
Db 13422 CGCAACGCCGCCAACCGGAGAGCGGCCCGCATCGAGTGTGAGCAAGGCCCTTCTCC 13481  
Qy 101 ArgLeuLysThrThrLeuProThrValProProAspThrLysLeuSerLysLeuAspThr 120  
Db 13482 AGACTCAAGACCACTCCCTGCGCTGGTGGCCCGCCGACACCAAGCTCTCCAAAGTGGACAG 13541  
Qy 121 LeuArgLeuAlaSerSerTrilleAlaHisLeuArgGlnIleLeuAlaAsnAspLysTyr 140  
Db 13542 CTAGCTGGCGTCCAGCTACATCGCCACTTGAGCAGATCTTGCTTAACGACAAATAC 13601  
Qy 141 GluAsnGlyTyrIleHisProValAsnLeu 150  
Db 13602 GAGAACGGGTACATTCAACCGGTCAACCTG 13631

RESULT 6  
LOCUS AF047418 1202 bp mRNA linear ROD 06-OCT-1998  
DEFINITION Mus musculus epicardin mRNA, complete cds.  
ACCESSION AF047418  
VERSION AF047418.1 GI:2921850  
KEYWORDS  
SOURCE Mus musculus (house mouse)  
ORGANISM

REFERENCE  
1 (bases 1 to 1202)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
AUTHORS Robb L., Mifsud, L., Hartley, L., Biben, C., Copeland, N.G.,  
Gilbert, D.J., Jenkins, N.A. and Harvey, R.P.

TITLE  
epicardin: A novel basic helix-loop-helix transcription factor gene  
expressed in epicardium, branchial arch myoblasts, and mesenchyme  
of developing lung, gut, kidney, and gonads  
Dev. Dyn. 213 (1), 105-113 (1998)

JOURNAL  
MEDLINE 98401941  
PUBMED 9733105  
REFERENCE 2 (bases 1 to 1202)  
AUTHORS Robb L. and Mifsud, L.  
TITLE Direct submission  
JOURNAL Submitted (10-FEB-1998) Haematology and Cancer, Walter and Eliza

FEATURES  
source

Hall Institute of Medical Research, PO Royal Melbourne Hospital,  
Melbourne, VIC 3050, Australia  
Location/Qualifiers  
1. .1202  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="Balb/c"  
/db\_xref="taxon:10090"  
199. 738  
/function="basic helix-loop-helix transcription factor"  
/notes="expressed in embryonic epicardium and  
mesenchyme-derived tissues of lung, gut, kidney, and  
gonad"  
/codon\_start=1  
/product="epicardin"  
/protein\_id="AAC62531.1"  
/db\_xref="GI:2921851"  
/translation="MSTGSLSDYEDLQEVMLDCDSLKVDNKEFCTSNESTEGSGNC  
ENSGPQKRGGLGRRKAPTKPSLGSVQEGKQVORNANARERMRMVLKAFSL  
KTTLPWVPPTKLSKLDLFLASSYLAHRLQILANDKYENGYIHPVNLTPFWFVAGKP  
ENDLKEYVTANRLCGTTAS"

## CDS

## ORIGIN

Alignment Scores:  
Pred. No.: 5.99e-120 Length: 1202  
Score: 120.00 Matches: 120  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 67.04% Indels: 0  
DB: 10 Gaps: 0

US-09-701-674A-23 (1-179) x AF047418 (1-1202)

Qy 42 SerAsnCysGluAsnGlySerProGlnLysGlyArgGlyLeuGlyLysArgArgLys 61  
Db 322 TCCAACTGCGAGAACGGGTCTCCACAGAGGGTCCGGGTGGCTAGGCAAGAGGAGAG 381  
Qy 62 AlaProThrLysLysSerProLeuSerGlyValSerGlnGluGlyLysGlnValGlnArg 81  
Db 382 GCGCCCACTAAGAAAACCCCGCTCAGCGGGGTGAGCCAGGAGGCAAGAGTCCACGCG 441  
Qy 82 AsnAlaAlaAsnAlaArgGluArgAlaArgMetArgValLeuSerLysAlaPheSerArg 101  
Db 442 AACCGGCCAATGCTGTGAGCGGCCCGCATCGGGTGTGAGCAAGGCTTCTCCAGG 501  
Qy 102 LeuLysThrThrLeuProThrValProProAspThrLysLeuSerLysLeuAspThrLeu 121  
Db 502 CTCAGACCAACCCCTGCGCTGGTCCCGCCGACCAAGCTCTCCAAGCTGGACACTCTC 561  
Qy 122 ArgLeuAlaSerSerTrilleAlaHisLeuArgGlnIleLeuAlaAsnAspLysTyrGlu 141  
Db 562 AGGCTGGGTCCAGTACATCGCTCACTTAAGCAGATCTTGGCCACGACAGTACGAG 621  
Qy 142 AsnGlyTyrIleHisProValAsnLeuThrTrpProPheMetValAlaGlyLysProGlu 161  
Db 622 AACGGTTACATTCAACCCAGTCAACCTGACGTGGCCCTTTATGTGGCCGCAACACGAG 681

## RESULT 7

AF029753

LOCUS AF029753 1217 bp mRNA linear ROD 25-FEB-2000  
DEFINITION Mus musculus basic helix-loop-helix factor Cori mRNA, complete cds.  
ACCESSION AF029753  
VERSION AF029753.1 GI:2613018  
KEYWORDS  
SOURCE Mus musculus (house mouse)  
ORGANISM

REFERENCE  
1 (bases 1 to 1217)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
AUTHORS Hidai, H., Bardales, R., Goodwin, R., Quettermous, T. and  
Quettermous, E.E.  
TITLE  
Cloning of capulin, a basic helix-loop-helix factor expressed in  
progenitor cells of the pericardium and the coronary arteries

Mech. Dev. 73 (1), 33-43 (1998) April 102(a)

JOURNAL  
MEDLINE  
PUBMED  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
FEATURES  
source

98213629  
9545526  
2 (bases 1 to 1217)  
Quettermous, T., Quertermous, E.E., Hidaï, H., Fadel, B., Boutet, S.C. and Bardales, R.  
Direct Submission  
Submitted (10-OCT-1997) Department of Medicine, Division of Cardiology, Stanford University, Falk Bldg, 300 Pasteur Drive, Stanford, CA 94305, USA  
Location/Qualifiers  
1..1217  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
/tissue\_type="pericardium and coronary arteries"  
1..1217  
/gene="Corl"  
213..752  
/gene="Corl"  
/function="transcription factor"  
/codon\_start=1  
/product="basic helix-loop-helix factor Corl"  
/protein\_id="AAB84256.1"  
/db\_xref="GI:2613019"  
/translation="MTSGLSLSDVEDLQEVEMLCDSLKVDNSKEFQTSNESTEEGNC  
ENGSPQGGGKGGKRRKATKKPSLGSVQEGKQVORNAAFRAMRVLGSKAFRL  
KTLPLVPPTDKLSKLDLTLASSYIAHLRIQLANDKYENGYIHPVNLTPFWVAGK  
ENDLKRVWTVANRLCGTTAS"

## ORIGIN

Alignment Scores:		
Pred. No.:	Length:	1217
Score:	120.00	Matches: 120
Percent Similarity:	100.00%	Conservative: 0
Best Local Similarity:	100.00%	Mismatches: 0
Query Match:	67.04%	Indels: 0
DB:	10	Gaps: 0
US-09-701-674A-23 (1-179) x AF029753 (1-1217)		
QY	42	SerAsnCYsGIuAsnGLySerProGlnLYsGLyArgGLyLeuGLyLYsArgArgLYs 61
Db	336	TCCAACCTCGAGAAAGCGGTCTCCACAGAGCGGTGCGGTGCGCTAGCGAAGAGGAGGAAG 395
QY	62	AlaProThrLYsLYsSerProLeuSerGLyValSerGlnGLyLYsGlnValGlnArg 81
Db	396	GGCGGCCACTAAGAAAAGCCGCTCAGCGGGGTGAGCCAGGAGGCGAAGCAGGTCCAGGCGC 455
QY	82	AsnAlaAlaAsnAlaATgGluArgAlaArgMetArgValLeuSerLYsAlaPheSerArg 101
Db	456	AACGGGGCCAAATGCTCGTGACGGGGCCCGCATCGGGTGTGTAGCAAGAGCCCTTCTCCAGG 515
QY	102	LeuLYsThrThrLeuProTyrValProProAspThrLYsLeuSerLYsLeuAspThrLeu 121
Db	516	CTCAAGACCAACCTGCGCCTGGGTGGTCCCGCGACCAACCAAGCTTCTCCAAGGTGGACACTCTC 575
QY	122	ArgLeuAlaSerSerTyrIleAlaHisLeuArgGlnIleLeuAlaAsnAspLYsTyrGlu 141
Db	576	AGSGTGGGGTCCAGCTACATCGCTACATTTAAGCAGAGATCTCTGGCCCAACGACCAAGTACGAG 635
QY	142	AsnGLyTyrIleHisProValAsnLeuThrTyrProPheMetValalaGLyLYsProGlu 161
Db	636	AACGGTGTACATTACCCAGTCAACCTGACGTGGGCCCTTTATGTTGGCGCGCAAAACACAG 695

RESULT	8	1220 bp	linear	ROD 11-DEC-1997
LOCUS	AB009453			
DEFINITION	Mus musculus mRNA for neptgonadin,			
ACCESSION	AB009453			
VERSION	AB009453.1			
KEYWORDS	neptgonadin.			
SOURCE	Mus musculus (house mouse)			

1025

**ORGANISM**  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 1220)  
**TITLE**  
Identification of Nephgonadin, a novel basic helix-loop-helix gene  
**JOURNAL**  
Published only in DataBase (1997)  
**REFERENCE**  
Tamura,M. and Nakatsuji,N.  
**AUTHORS**  
Tamura,M. and Nakatsuji,N.  
**TITLE**  
Direct Submission  
**JOURNAL**  
Submitted (04-DEC-1997) Masaru Tamura, National Institute of  
Genetics, Mammalian Development; Yata 1111, Mishima, Shizuoka 411,  
Japan (E-mail:matamur@lab.nig.ac.jp, Tel:+81-559-81-6832,  
Fax:+81-559-81-6828)  
**FEATURES**  
Location/Qualifiers  
1..1220  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
223..762  
/function="transcription factor"  
/note="basic helix-loop-helix gene"  
/codon\_start=1  
/product="nephgonadin"  
/protein\_id="BAA23883.1"  
/db\_xref="GI:2696118"  
/translation="MSTGSLSDVEDLQEVMLDCSLKVDSEKFGTSTNEEGSNC  
ENGSPKRGKGGKRRKATPKKSLGVSQEGQVQRNNAARERAMRVLSKAFRSL  
KTLPLWPVPTKLSDLTLDLASSYIAHLRQILANDKYENGYIHPVNLTFPFWVAGKPR  
ENDLKEVTVANRLCGTAS"  
1220  
polyA\_site  
/note="17 A nucleotides"  
**ORIGIN**  
Alignment Scores:  
Pred. No.: 6,07e-120 Length: 1220  
Score: 120.00 Matches: 120  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 67.04% Indels: 0  
DB: 10 Gaps: 0  
US-09-701-674A-23 (1-179) x AB009453 (1-1220)  
QY 42 SerAsnCyseGluAsnGlySerProGlnLysGlyValGlyGlyLeuGlyLysArgArgLys 61  
Db 346 TCCAACTGCGAGAACGGGTCTCCACAGAGGGTCCGGGTGGCTAGCAAGAGGAGGAG 405  
QY 62 AlaProThrLysLysSerProLeuSerGlyValSerGlnGluGlyLysGlnValGlnArg 81  
Db 406 GCGCCCACTAAGAAAAGCCCGTCTACGCGGGTCCAGCAGAGGGCAAGCAAGTCCACGCG 465  
QY 82 AsnAlaAlaAsnAlaArgGluArgAlaArgMetArgValLeuSerLysAlaPheSerArg 101  
Db 466 AACCGCGCCATGCTCGTAGCGGGCCCGCATCCGGGTGCTGAGCAAGGCGCTTCTCCAGG 525  
QY 102 LeuLysThrThrLeuProTrpValProProAspThrLysLeuSerLysLeuAspThrLeu 121  
Db 526 CTCAGACCAACCTCGCTGGGTGGCCCGGACACCAAGCTCTCCAAGCTGGACACTCTC 585  
QY 122 ArgLeuAlaSerSerTyrIleAlaHisLeuArgGlnIleLeuAlaAsnAspLysTyrGlu 141  
Db 586 AGGCTGGCTCCAGCTACATCGCTCACTTAAGGCAGATCTCTGGCCCAACGCAAGTAGCAG 645  
QY 142 AsnGlyTyrIleHisProValAsnLeuThrTrpProPheMetValAlaGlyLysProGlu 161  
Db 646 AACGGTTACATTCAACCCAGTCAACCTGACGTGGCGCCCTTTATGTGTGGCGGCAACACGAG 705  
**RESULT 9**  
AF035717  
LOCUS  
DEFINITION  
Mus musculus mesoderm-specific basic-helix-loop-helix protein  
(Pod1) mRNA, complete cds.  
AF035717 1240 bp mRNA linear ROD 05-OCT-1998

```

ACCESSION AF035717
VERSION AF035717.1 GI:2745884
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 1240)
AUTHORS Quaglin, S.E., Vanden Heuvel, G.B. and Igarashi, P.
TITLE Pod-1, a mesoderm-specific basic-helix-loop-helix protein expressed in mesenchymal and glomerular epithelial cells in the developing kidney. Dev. 71 (1-2), 37-48 (1998) Feb 16 2004
JOURNAL Mech. Dev. 71 (1-2), 37-48 (1998)
MEDLINE 98175875
PUBMED 9507058
REFERENCE 2 (bases 1 to 1240)
AUTHORS Quaglin, S.E., Vanden Heuvel, G.B. and Igarashi, P.
TITLE Direct Submission
JOURNAL Submitted (24-NOV-1997) Internal Medicine, Yale University, 333 Cedar Street, New Haven, CT 06520-8029, USA
FEATURES
     source
       1. 1240
         /organism="Mus musculus"
         /mol_type="mRNA"
         /db_xref="taxon:10090"
         /chromosome="10"
         /map="near D10Mit2"
         /genes="Pod1"
         /gene="Pod1"
         /note="Pod-1"
         /codon_start=1
         /product="mesoderm-specific basic-helix-loop-helix protein"
         /protein_id="AAC62513.1"
         /db_xref="GI:2745885"
         /translation="MSGTGLSDVDLQVEMDCDLSKVDNKGFTGNESTEGSNC
         ENGSPKGGGLGKPKRKATKPSGVSGVSGKQVORNAARERARVLSKAPSL
         KTLWPVPTKLSKLDLTLARLASSVIAHLRLQLANDKYENGVIHFNLTWTFPMVAGKP
         ENDLKSVTVANRLCGTTAS"
     misc_feature
       457..618
         /gene="Pod1"
         /note="encodes basic-helix-loop-helix domain"
ORIGIN
Alignment Scores:
 6.17e-120 Length: 1240
Score: 120.00 Matches: 120
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 67.04% Indels: 0
DB: 10 Gaps: 0
US-09-701-674A-23 (1-179) x AF035717 (1-1240)
QY 42 SerAsnCysGluAsnGlySerProGlnLysGlyArgGlyLeuGlyLysArgLys 61
DB 349 TCCAACTGGAGAACGGGCTCCACAGAGAGGTGGGCTAGGCAAGAGGAGGAAG 408
QY 62 AlaProThrLysSerProLeuSerGlyValSerGlnGlyLysGlnValGlnArg 81
DB 409 GCGCCCACTAAGAAACCCCGTTCACGGGGGTACGACGAGGGGCAAGCAGTCCAGGCG 469
QY 82 AsnAlaAsnAlaArgGluArgAlaArgMetArgValLeuSerLysAlaPheSerArg 101
DB 469 AACCGCGGCAATGCTGTCGCGGGGCCCATGCGGGTCTGACGACAGGCTTCTCCAGG 528
QY 102 LeuLysThrThrLeuProTrpValProProAspThrLysLeuSerLysLeuAspThrLeu 121
DB 529 CTCAGACCAACCCCTGCGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 588
QY 122 ArgLeuAlaSerSerTyrIleAlaHisLeuArgGlnIleLeuAlaAsnAspLysTyrGlu 141

```

```

DB 589 AGGCTGGCGTCCAGCTACATCGTCACTTAAGGCAGATCTGCGCAACGACAGTACGAG 648
QY 142 AsnGlyTyrIleHisProValAsnLeuThrTyrProPheMetValAlaGlyLysProGlu 161
DB 649 AACGTTACATTCACCCAGTCAACCTGACGTGCGCCCTTTATGTTGGCCGCAACAGAG 708
RESULT 10
LOCUS BC053525
DEFINITION BC053525 1246 bp mRNA linear ROD 30-JUN-2004
ACCESSION BC053525
VERSION BC053525.1 GI:31566109
SOURCE MGC.
ORGANISM Mus musculus (house mouse)
REFERENCE 1 (bases 1 to 1246)
AUTHORS Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
1 (bases 1 to 1246)
Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,
Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,
Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F.,
Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,
Stapleton, M., Soares, M.B., Bonaldo, M.P., Casavant, T.L.,
Schaefer, T.E., Brownstein, M.J., Usdin, T.B., Ioshizuka, S.,
Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J.,
Abramson, R.D., Mullaly, S.J., Bosak, S.A., McEwan, P.J.,
McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,
Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,
Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,
Fahey, J., Helton, E., Kettner, M., Madan, A., Rodriguez, S.,
Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y.,
Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,
Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,
Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smalys, D.E.,
Schnerch, A., Schein, J.E., Jones, S.J. and Marra, M.A.
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
12477932
2 (bases 1 to 1246)
Strausberg, R.
Direct Submission
Submitted (09-JUN-2003) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: gqabs-x@mail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: National Institutes of Health Intramural
Sequencing Center (NISC),
Gaithersburg, Maryland;
Web site: http://www.nisc.nih.gov/
Contact: nisc_gnc@hgr.nih.gov
Ahter, N., Ayele, K., Beckstrom-Sternberg, S.M., Benjamin, B.,
Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S.,
Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghighi, P.,
Hansen, N., Ho, S.-L., Karlins, E., Kwong, P., Laric, P., Legaspi, R.,
Maduro, Q.L., Masiello, C., Maskeri, B., Mastrian, S.D., McCloskey, J.C.,
McDowell, J., Pearson, R., Stantripop, S., Thomas, P.J., Touchman, J.W.,
Turgeon, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L.,
Young, A., Zhang, L.-H. and Green, E.D.
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAC Plate: 108 Row: p Column: 16.
Location/Qualifiers

```

```

source      1. .1246
/mol_type="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="MGC:58914 IMAGE:6529241"
/tissue_type="Colon, normal. 5 month old male mouse."
/clone_lib="NCI_CGAP_C024"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"
gene        1. .1246
/genes="Tcf21"
/note="synonyms: epc, Pod1, capsulin, epicardin, Pod-1"
/db_xref="LocusID:21412"
/db_xref="MGI:1202715"
233. .772
/genes="Tcf21"
/codon_start=1
/product="transcription factor 21"
/protein_id="AAH53525.1"
/db_xref="GI:31566110"
/db_xref="LocusID:21412"
/db_xref="MGI:1202715"
/translation="MSTGSLSDVEDLQEVEMLDCLSKVDSNKEFGTSTNESTEGSNC
ENGSPKGRGLGKRRKAPTCKSPKSGVQEGKQVORNAANARERARMRVLSKAFSRL
KTTLPVPDPTKLSKLDLRLASSYIAHLRLQILANDKYENGVIHPVNLTPWPMVAGKP
ENDLKEVVTANLGGTTAS"
ORIGIN
Alignment Scores:      6.2e-120      Length:      1246
Pred. No.:      120.00      Matches:      120
Score:      100.00%      Conservative:      0
Percent Similarity:      100.00%      Mismatches:      0
Best Local Similarity:      100.00%      Indels:      0
Query Match:      67.04%      Gaps:      0
DB:      10
US-09-701-674A-23 (1-179) x BC053525 (1-1246)
QY      42 SerAsnCysGluAsnGlySerProGlnLysGlyValSerGlyGlyLeuGlyLysArgArgLys 61
Db      356 TCCAACTGGAGAACGGGTCTCCACAGAGGGTCCGGTGGCTAGGCAAGAGGAGGAG 415
QY      62 AlaProThrLysLysSerProLeuSerGlyValSerGlnGlnGlyLysGlnValGlnArg 81
Db      416 GCGCCCACTAAGAAAAGCCGCTCAGCGGGTCAGCCAGGCGGCAAGCAGGTCCAGCGC 475
QY      82 AsnAlaAlaAsnAlaArgGluArgAlaArgMetArgValLeuSerLysAlaPheSerArg 101
Db      476 AACCGCGCCAATGCTCGTAGCGGGCCCGCATGCGGGTCTGAGCAAGGCCCTCTCCAGG 535
QY      102 LeuLysThrThrLeuProTyrValProProAspThrLysLysLysLeuAspThrLeu 121
Db      536 CTCAGACCAACCTCGCTGGTGGTCCCGCGACCAAGCTCTCCAGCTGGACACTCTC 595
QY      122 ArgLeuAlaSerSerTyrIleAlaHisLeuArgGlnIleLeuAlaAsnAspLysTyrGlu 141
Db      596 AGGCTGGCTCCAGCTACATCGCTCACTTAAGCAGATCTCTGGCAACAGCAAGTACGAG 655
QY      142 AsnGlyTyrIleHisProValAsnLeuThrTyrProPheMetValAlaGlyLysProGlu 161
Db      656 AACGGTTACATTCACCCAGTCAACCTGACGTGGCCCTTTATGGTGGCCGCAACACGAG 715
RESULT 11
AF036945      1267 bp      mRNA      linear      ROD 15-JUN-1998
LOCUS      Mus musculus capsulin mRNA, complete cds.
DEFINITION      AF036945
ACCESSION      AF036945.1 GI:2695694
VERSION      AF036945.1
KEYWORDS      Mus musculus (house mouse)
SOURCE      Mus musculus
ORGANISM      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

```

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (Bases 1 to 1267)
Lu,J., Richardson,J.A. and Olson,E.N.
Capsulin: a novel bHLH transcription factor expressed in epicardial
progenitors and mesenchyme of visceral organs
Mech. Dev. 73 (1), 23-32 (1998) Apr 002 (a)
98213628
PUBMED
MEDLINE
REFERENCE 2 (Bases 1 to 1267)
AUTHORS      Lu,J. and Olson,E.N.
TITLE      Direct Submission
JOURNAL      Submitted (04-DEC-1997) Molecular Biology and Oncology, UT
Southwestern Medical Center, 6000 Harry Hines Blvd., NA8.510,
Dallas, TX 75235-9148, USA
FEATURES
source      1. .1267
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
264. .803
/function="BHLH transcription factor"
/codon_start=1
/product="capsulin"
/protein_id="AAC23537.1"
/db_xref="GI:2695695"
/translation="MSTGSLSDVEDLQEVEMLDCLSKVDSNKEFGTSTNESTEGSNC
ENGSPKGRGLGKRRKAPTCKSPKSGVQEGKQVORNAANARERARMRVLSKAFSRL
KTTLPVPDPTKLSKLDLRLASSYIAHLRLQILANDKYENGVIHPVNLTPWPMVAGRP
ENDLKEVVTANLGGTTAS"
ORIGIN
Alignment Scores:      6.3e-120      Length:      1267
Pred. No.:      120.00      Matches:      120
Score:      100.00%      Conservative:      0
Percent Similarity:      100.00%      Mismatches:      0
Best Local Similarity:      100.00%      Indels:      0
Query Match:      67.04%      Gaps:      0
DB:      10
US-09-701-674A-23 (1-179) x AF036945 (1-1267)
QY      42 SerAsnCysGluAsnGlySerProGlnLysGlyValSerGlyGlyLeuGlyLysArgArgLys 61
Db      387 TCCAACTGGAGAACGGGTCTCCACAGAGGGTCCGGTGGCTAGGCAAGAGGAGGAG 446
QY      62 AlaProThrLysLysSerProLeuSerGlyValSerGlnGlnGlyLysGlnValGlnArg 81
Db      447 GCGCCCACTAAGAAAAGCCGCTCAGCGGGTCTAGCCAGGCGGCAAGGCTCCAGCGC 506
QY      82 AsnAlaAlaAsnAlaArgGluArgAlaArgMetArgValLeuSerLysAlaPheSerArg 101
Db      507 AACCGCGCCAATGCTCGTAGCGGGCCCGCATGCGGGTCTGAGCAAGGCCCTCTCCAGG 566
QY      102 LeuLysThrThrLeuProTyrValProProAspThrLysLysLysLeuAspThrLeu 121
Db      567 CTCAGACCAACCTCGCTGGTGGTCCCGCGACCAAGCTCTCCAAGCTGGACACTCTC 626
QY      122 ArgLeuAlaSerSerTyrIleAlaHisLeuArgGlnIleLeuAlaAsnAspLysTyrGlu 141
Db      627 AGGCTGGCTCCAGCTACATCGCTCACTTAAGCAGATCTCTGGCAACAGCAAGTACGAG 686
QY      142 AsnGlyTyrIleHisProValAsnLeuThrTyrProPheMetValAlaGlyLysProGlu 161
Db      687 AACGGTTACATTCACCCAGTCAACCTGACGTGGCCCTTTATGGTGGCCGCAACACGAG 746
RESULT 12
AF035718      1254 bp      mRNA      linear      PRI 05-OCT-1998
LOCUS      Homo sapiens mesoderm-specific basic-helix-loop-helix protein
DEFINITION      (POD1) mRNA, complete cds.
ACCESSION      AF035718
VERSION      AF035718.1 GI:2745886
KEYWORDS

```



Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,  
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,  
Zembek,L., Zimmer,A. and Zody,M.

# TITLE

Submitted (21-AUG-2002) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Aug 21, 2002 this sequence version replaced gi:17060067.  
All repeats were identified using RepeatMasker:  
Smith, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html

## COMMENT

----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence\_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L16369

Center Clone name: 103.E.4

----- Summary Statistics

Sequencing vector: Plasmid; n/a; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 178139 bases at least Q40

Consensus quality: 182376 bases at least Q30

Consensus quality: 183759 bases at least Q20

Insert size: 195000; agarose-fp

Insert size: 184732; sum-of-contigs

Quality coverage: 6.4 in Q20 bases; agarose-fp

Quality coverage: 6.8 in Q20 bases; sum-of-contigs

-----

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 27 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 23755: contig of 23755 bp in length  
\* 23756: gap of 100 bp  
\* 23856: contig of 674 bp in length  
\* 24529: gap of 100 bp  
\* 24630: contig of 1166 bp in length  
\* 25795: gap of 100 bp  
\* 25796: contig of 1079 bp in length  
\* 25896: gap of 100 bp  
\* 26974: gap of 100 bp  
\* 26975: contig of 1530 bp in length  
\* 27074: gap of 100 bp  
\* 27075: contig of 100 bp  
\* 28004: gap of 100 bp  
\* 28005: contig of 1363 bp in length  
\* 30068: gap of 100 bp  
\* 30168: contig of 921 bp in length  
\* 31089: gap of 100 bp  
\* 31189: contig of 1770 bp in length  
\* 32958: gap of 100 bp  
\* 33059: contig of 2172 bp in length  
\* 35231: gap of 100 bp  
\* 35331: contig of 2077 bp in length  
\* 37408: gap of 100 bp  
\* 37508: contig of 2695 bp in length  
\* 40202: gap of 100 bp  
\* 40302: contig of 3531 bp in length  
\* 40303: gap of 100 bp  
\* 43834: contig of 2423 bp in length  
\* 43934: gap of 100 bp  
\* 46356: contig of 2185 bp in length  
\* 46457: gap of 100 bp  
\* 48641: contig of 2939 bp in length  
\* 48741: gap of 100 bp  
\* 48742: contig of 3211 bp in length  
\* 51681: gap of 100 bp  
\* 51781: contig of 100 bp  
\* 54991: gap of 100 bp  
\* 55091: contig of 3487 bp in length  
\* 55092: gap of 100 bp  
\* 58578: contig of 3506 bp in length  
\* 58579: gap of 100 bp  
\* 62184: contig of 3506 bp in length

Alignment Scores:

\* 62185 62284: gap of 100 bp  
\* 62285 69983: contig of 7699 bp in length  
\* 69984 70083: gap of 100 bp  
\* 70094 79364: contig of 9281 bp in length  
\* 79365 79464: gap of 100 bp  
\* 79465 87765: contig of 8301 bp in length  
\* 87766 87865: gap of 100 bp  
\* 87866 97785: contig of 9920 bp in length  
\* 97786 97886: gap of 100 bp  
\* 97887 130019: contig of 32134 bp in length  
\* 97888 130119: gap of 100 bp  
\* 130020 144101: contig of 13982 bp in length  
\* 144102 144201: gap of 100 bp  
\* 144202 158508: contig of 14307 bp in length  
\* 158509 158608: gap of 100 bp  
\* 158609 174086: contig of 15478 bp in length  
\* 174087 174186: gap of 100 bp  
\* 174187 187332: contig of 13146 bp in length.  
\* Location/Qualifiers  
\* 1..187332  
\* /organism="Mus musculus"  
\* /mol\_type="genomic DNA"  
\* /db\_xref="taxon:10090"  
\* /clone\_lib="RPCI-23 Female Mouse BAC"  
\* 1..23755  
\* /note="assembly\_fragment  
\* clone\_end:SP6  
\* vector\_side:left"  
\* 23856..24529  
\* /note="assembly\_fragment"  
\* 24630..25795  
\* /note="assembly\_fragment"  
\* 25896..26974  
\* /note="assembly\_fragment"  
\* 27075..28604  
\* /note="assembly\_fragment"  
\* 28705..30067  
\* /note="assembly\_fragment"  
\* 30168..31088  
\* /note="assembly\_fragment"  
\* 31189..32958  
\* /note="assembly\_fragment"  
\* 33059..35230  
\* /note="assembly\_fragment"  
\* 35331..37407  
\* /note="assembly\_fragment"  
\* 37508..40202  
\* /note="assembly\_fragment"  
\* 40303..43833  
\* /note="assembly\_fragment"  
\* 43934..46356  
\* /note="assembly\_fragment"  
\* 46457..48641  
\* /note="assembly\_fragment"  
\* 48742..51680  
\* /note="assembly\_fragment"  
\* 51781..54991  
\* /note="assembly\_fragment"  
\* 55092..58578  
\* /note="assembly\_fragment"  
\* 58679..62184  
\* /note="assembly\_fragment"  
\* 62285..69983  
\* /note="assembly\_fragment"  
\* 70084..79364  
\* /note="assembly\_fragment"  
\* 79465..87765  
\* /note="assembly\_fragment"  
\* 87866..97785  
\* /note="assembly\_fragment"



```

Pred. No.:      6.32e-106      Length:      187332
Score:          109.00      Matches:      109
Percent Similarity: 100.00%      Conservatives: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:      60.89%      Indels: 0
DB:              2      Gaps: 0

US-09-701-674A-23 (1-179) x AC101292 (1-187332)

QY      42 SerAsnCysGluAsnGlySerProGlnLysGlyArgGlyGlyLeuGlyLysArgArgLys 61
Db      82639 TCCAACTGGCAGAACGGGTCTCCACAGAAGGGTGGCGGTGGCTAGGCAAGAGGAGGAG 82580

QY      62 AlaProThrLysLysSerProLeuSerGlyValSerGlnGluGlyLysGlnValGlnArg 81
Db      82579 GCGCCCACTAAGAAAGCCGCTCAGCGGGGTGTCAGCAGAGGGGCAAGGAGTCCAGGCG 82520

QY      82 AsnAlaAlaAsnAlaArgGluArgAlaArgMetArgValLeuSerLysAlaPheSerArg 101
Db      82519 AACGCGGCCAAATGCTGTCAGCGGGGCCGTCATGCGGGTGTCTGAGCAAGGCTTCTCCAGG 82460

QY      102 LeuLysThrThrLeuProThrValProProAspThrLysLeuSerLysLeuAspThrLeu 121
Db      82459 CTCAGAGACCACTGCGCTGGGTGGTGGCCCGGACCCAGCTCTCCAGTGGACACTTCTC 82400

QY      122 ArgLeuAlaSerSerTyrlleAlaHisLeuArgGlnIleLeuAlaAsnAspLysTyrglu 141
Db      82399 AGGCTGGGTGTCAGGTACATGCTACTTAAGCAGATCTCGGCCAACGACAGTACGAG 82340

QY      142 AsnGlyTyrlleHisProValAsnLeu 150
Db      82339 AACGGTTACATTCACCCAGTCAACCTG 82313

RESULT 14
LOCUS   AC115183      241544 bp      DNA      linear      HTG 20-NOV-2002
DEFINITION Rattus norvegicus clone CH230-58J14, WORKING DRAFT SEQUENCE, 3
unorderd pieces.
ACCESSION AC115183
VERSION   HTG; HTGS PHASE1; HTGS DRAFT; HTGS FULLTOP.
KEYWORDS  Rattus norvegicus (Norway rat)
SOURCE    Rattus norvegicus
ORGANISM  Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 241544)
AUTHORS   Muzny,D,Marie., Metzker,M.Lee., Abramson,S., Adams,C., Alder,J.,
Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
Egan,A., Escoto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
Gebelegorgis,E., Geer,K., Gill,R., Grady,M., Guerra,M., Guevara,W.,
Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
Harvey,X., Haylak,P., Hawes,A., Henderson,N., Hernandez,J.,
Herrandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
Karpachy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
Kowls,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
Lorenshewa,L., Loulsegged,H., Lozado,R.J., Lu,X., Ma,J.,
Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,

```

```

Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
Mawhney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
Narkervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
Nwackemeleh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkuch,C.,
Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
Puzo,M., Quirz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
Rives,C., Rocky,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
Sandes,J., Savary,G., Schefer,S., Scott,G., Shatsman,S., Shen,H.,
Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sooa,J.,
Steinle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Umaní,K.,
Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
Weinstock,G. and Gibbs,R.A.
Direct Submission
Unpublished
2 (bases 1 to 241544)
Worley,K.C.
Direct Submission
Submitted (15-MAR-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 241544)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (20-NOV-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Nov 20, 2002 this sequence version replaced gi:22856693.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GSCC
Center clone name: CH230-58J14
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 214521 bases at least Q40
Consensus quality: 217222 bases at least Q30
Consensus quality: 219223 bases at least Q20
Estimated insert size: 218736; sum-of-contigs estimation
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
consists of 3 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.

```

\* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

1 232705: contig of 232705 bp in length  
 \* 232706 232805: gap of unknown length  
 \* 232806 234023: contig of 1218 bp in length  
 \* 234024 234123: gap of unknown length  
 \* 234124 241544: contig of 7421 bp in length.

# FEATURES

Location/Qualifiers  
 1..241544  
 /organism="Rattus norvegicus"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:10116"  
 /clone="CH230-58U14"  
 1..1433  
 /note="wgs\_contig"

## ORIGIN

Alignment Scores:  
 Pred. No.: 8,05e-106 Length: 241544  
 Score: 109.00 Matches: 109  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 60.89% Indels: 0  
 DB: 2 Gaps: 0

US-09-701-674A-23 (1-179) x AC115183 (1-241544)

QY 42 SerAsnCysGluAsnGlySerProGlnLysGlyArgGlyGlyLeuGlyLysArgGlyLys 61  
 Db 76907 TCCAACTGCGAGAACGGGTCTCCACAGAAAGGTGCTGGTGGCTGGCGAAGAGGAGGAG 76848  
 QY 62 AlaProThrLysSerProLeuSerGlyValSerGlnGlnGlyLysGlnValGlnArg 81  
 Db 76847 GCGCCACCAAGAAAGCCGTCAGCGGGGTGAGCCAGGAGGGGCAAGCAGGTCCAGCGC 76788  
 QY 82 AsnAlaAlaAsnAlaArgGluArgAlaArgMetArgValLeuSerLysAlaPheSerArg 101  
 Db 76787 AACCGGGCCAAATGCACGCGAGCGGGCCGCGATGCGGGTCTGAGCAAGGCTTCTCCAGG 76728  
 QY 102 LeuLysThrThrLeuProTrpValProProAspThrLysLeuSerLysLeuAspThrLeu 121  
 Db 76727 CTCAGACTACCTTCCCTGGGTGGTCCCGGACACCAAGCTCTCAAGCTGGACACTCTC 76668  
 QY 122 ArgLeuAlaSerSerTyrlleAlaHisLeuArgGlnIleLeuAlaAsnAspLysTyrlle 141  
 Db 76667 AGGCTGGCTCCAGCTACATCGCGACCTGAGCGAGATCTTGGCCACACGACAGTACGAG 76608  
 QY 142 AsnGlyTyrlleHisProValAsnLeu 150  
 Db 76607 AATGGTTACATTCAACCCAGTCAACCTG 76581

## RESULT 15

EX932800  
 LOCUS  
 DEFINITION  
 Gallus gallus finished cdna, clone CHEST265h24.  
 ACCESSION  
 EX932800  
 VERSION  
 EX932800.2 GI:46017571  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Gallus gallus (chicken)

REFERENCE  
 AUTHORS  
 Boardman,P.E., Bonfield,J.K., Brown,W.R.A., Carder,C., Chalk,S.E.,  
 Croning,M.D.R., Davies,R.M., Francis,M.D., Graham,D.V.,  
 Hubbard,S.J., Humphray,S.J., Hunt,P.J., Maddison,M., McLaren,S.R.,  
 Niblett,D., Overton,I.M., Rogers,J., Scott,C.E., Taylor,R.G.,  
 Tickle,C. and Wilson,S.A.  
 Direct Submission

TITLE  
 JOURNAL  
 Submitted (29-MAR-2004) Sanger Institute, Hinxton, Cambridgeshire,  
 CE10 1SA, UK. E-mail enquiries: chickest@ms.unist.ac.uk

## COMMENT

On Apr 1, 2004 this sequence version replaced gi:41633328.  
 BBSRC/Dundee/Nottingham/Sanger/Sheffield/UMIST Gallus gallus cdna  
 sequencing project.  
 This sequence is from the  
 BBSRC/Dundee/Nottingham/Sanger/Sheffield/UMIST cdna collection,  
 from a library constructed by Elizabeth Bosch. cdna was prepared  
 from RNA extracted from ovary, normalised, and poly A-trimmed.  
 EcoRI-NotI cut cdna was then ligated into the vector. Vector:  
 pluescript II KS(+); Site\_1: EcoRI; Site\_2: NotI Host: Escherichia  
 coli DH10B.

## FEATURES

Location/Qualifiers  
 1..1164  
 /organism="Gallus gallus"  
 /mol\_type="mRNA"  
 /strain="Layer"  
 /db\_xref="taxon:9031"  
 /clone="CHEST265h24"  
 /clone\_lib="CSEORBN19"  
 /dev\_stage="adult"

## ORIGIN

Alignment Scores:  
 Pred. No.: 4,29e-96 Length: 1164  
 Score: 98.00 Matches: 98  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 54.75% Indels: 0  
 DB: 5 Gaps: 0

US-09-701-674A-23 (1-179) x BX932800 (1-1164)

QY 71 GlyValSerGlnGluGlyLysGlnValGlnArgAsnAlaAlaAlaGluArgAla 90  
 Db 377 GGTGTGAGCCAGGAGGAAAGAGGTCTCCAGAAACGCTCCACGACGAGGAGGAGGCA 436  
 QY 91 ArgMetArgValLeuSerLysAlaPheSerArgLeuLysThrThrLeuProTrpValPro 110  
 Db 437 AGGATGAGGTCTCTTAGCAAGCCTTCTCCAGGCTTAAGACCACTGCTGGTGGTGGCC 496  
 QY 111 ProAspThrLysLeuSerLysLeuAspThrLeuArgLeuAlaSerSerTyrlleAlaHis 130  
 Db 497 CCAGACACCAAGCTTTCACCACTGGACACCTTGGGTGGCTCCAGCTACATCGCACAC 556  
 QY 131 LeuArgGlnIleLeuAlaAsnAspLysTyrlleGluAsnGlyTyrlleHisProValAsnLeu 150  
 Db 557 CTGAGGCAATCTCTGGCCACGACAGTACGAAATGGCTACATCCATCCAGTCAACCTG 616  
 QY 151 ThrTrpProPheMetValAlaGlyLysProGluSerAspLeuLysGluValVal 168  
 Db 617 ACCTGGCCTTTTATGGTAGCGGCAACCCGAGAGTGCCTGAAAGAGTGGTG 670

Search completed: December 17, 2004, 02:26:58

Job time : 3891 secs

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: December 16, 2004, 20:19:00 ; Search time 3794 Seconds  
(without alignments)  
2231.117 Million cell updates/sec

Title: US-09-701-674A-23  
Perfect score: 917  
Sequence: 1 MTGSLSDVEDLQEVEMLEC.....PESDLKEVVTASRLCGTTAS 179

Scoring table: BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 4526729 seqs, 23644849745 residues

Total number of hits satisfying chosen parameters: 9053458

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame\_plus\_p2n.model -DRV=xlh  
-Q=/cgn2/USPTO.spool\_h/US09701674/runat\_15122004\_100440\_27505/app\_query.fasta\_1.327  
-DB=GenEmbl -QFMT=fastp -SUFFIX=rge -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USPR=US09701674 @CGN 1 1 3731 @runat\_15122004\_100440\_27505 -NCPU=6 -ICPU=3  
-NO\_WMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl.\*  
1: gb.ba.\*  
2: gb.htg.\*  
3: gb.in.\*  
4: gb.om.\*  
5: gb.ov.\*  
6: gb.pat.\*  
7: gb.ph.\*  
8: gb.pl.\*  
9: gb.pr.\*  
10: gb.ro.\*  
11: gb.sts.\*  
12: gb.sy.\*  
13: gb.un.\*  
14: gb.vi.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	917	100.0	537	9	CR450293 Homo sapi
2	917	100.0	1272	6	BD211771 Proteins
3	917	100.0	3231	9	BC025697 Homo sapi
4	914	99.7	1254	9	AF035718 Homo sapi

5	913	99.6	1257	9	AF047419 Homo sapi
6	904	98.6	1259	6	CQ718038 Sequence
7	887	96.7	1202	10	AF047418 Mus muscu
8	887	96.7	1217	10	AF029753 Mus muscu
9	887	96.7	1220	10	AB009453 Mus muscu
10	887	96.7	1240	10	AF035717 Mus muscu
11	887	96.7	1246	10	BC053525 Mus muscu
12	887	96.7	1267	10	AF036945 Mus muscu
13	839	91.5	1164	5	BX932800 Gallus ga
14	804	87.7	965	5	BC073597 Xenopus l
15	804	87.7	1059	5	AF660871 Xenopus l
16	766	83.5	25186	9	AL356109 Human DNA
17	748.5	81.6	187332	2	AC101292 Mus muscu
18	746	81.4	241544	2	AC115183 Rattus no
19	646	70.4	513	10	AF061752 Rattus no
20	568	61.9	124070	5	BX530074 Zebrafish
21	568	61.9	231207	2	CR293516 Danio rer
22	561	61.2	189465	2	CR318665 Danio rer
23	514	56.1	621	9	BT007169 Homo sapi
24	514	56.1	621	12	BT008287 Synthetic
25	514	56.1	1917	6	CQ717187 Sequence
26	514	56.1	2031	9	BC006313 Homo sapi
27	514	56.1	2040	9	BC067827 Homo sapi
28	507	55.3	2382	6	AX281674 Sequence
29	495.5	54.0	606	10	AF108216 Mus muscu
30	495.5	54.0	1718	10	AF087035 Mus muscu
31	493.5	53.8	1716	9	AF087036 Homo sapi
32	482	52.6	1893	9	AF060154 Homo sapi
33	398	43.4	170668	9	AC104012 Homo sapi
34	398	43.4	198331	9	AC022867 Homo sapi
35	388.5	42.4	163698	10	AC115876 Mus muscu
36	388.5	42.4	181157	10	AC121789 Mus muscu
37	388.5	42.4	225399	2	AC113255 Rattus no
38	362.5	39.5	222122	2	CR450806 Danio rer
39	353	38.5	240	6	AR270465 Sequence
40	318	34.7	697	6	AX333347 Sequence
41	318	34.7	697	6	AX333763 Sequence
42	266	29.0	144007	6	AC141743 Apis mell
43	212.5	23.2	1066	3	AD000729 Drosophila
44	211.5	23.1	426	6	CQ722685 Sequence
45	211.5	23.1	1066	6	CQ582567 Sequence

## ALIGNMENTS

CR450293 537 bp mRNA linear PRI 18-MAY-2004  
Homo sapiens full open reading frame cDNA clone RZPD0834B11D for  
gene TCF21, transcription factor 21; complete cds; without  
stopcodon.  
CR450293  
CR450293.1 GI:47496532  
Full ORF shuttle clone, Gateway(TM), complete cds.  
Homo sapiens (human)  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
Ebert,L., Schick,M., Neubert,P., Schatten,R., Henze,S. and Korn,B.  
Cloning of human full open reading frames in Gateway(TM) system  
entry vector (pDONR201)  
2 (bases 1 to 537)  
Ebert,L., Schick,M., Neubert,P., Schatten,R., Henze,S. and Korn,B.  
Direct Submission  
Submitted (18-MAY-2004) RZPD Deutsches Ressourcenzentrum fuer  
Genomforschung GmbH, Im Neuenheimer Feld 580, D-69120 Heidelberg,  
Germany  
RZPD: RZPD0834B11D, ORFNO 76  
www.rzpd.de/cgi-bin/products/cl.cgi?cloneID=RZPD0834B11D RZPDLFB:  
Human Full ORF Clones Gateway(TM) - RZPD (kan-resist.) RZPD LIB No.  
834

www.rzpd.de/cgi-bin/products/showLib.pl.cgi/response?libNo=834  
 www.rzpd.de/products/orfclones/  
 Contact: Ina Rolf  
 RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH,  
 Heubnerweg 6, D-14059 Berlin, Germany  
 Tel: +49 30 32639 101  
 Fax: +49 30 32639 111  
 www.rzpd.de

This clone is available from RZPD;  
 contact RZPD (customer.service@rzpd.de) for further information.  
 This CDS clone is a part of a collection of human full length  
 expression clones generated by RZPD.  
 This CDS has been cloned without stopcodon.  
 The CDS has been inserted into pDONR201 via a BP Clonase(TM)  
 reaction. Additional sequence has been added in front of the start  
 codon (ATG): att...AAAAA GGT GGC ACC CCT GGT CCA GGT (ATG)  
 After the last codon additional sequence has been added: CCA GGC  
 CCA GGC GGC G in front of the 3' att site (AC CCA GCT TTC TT).  
 Compared to the reference sequence NM\_003206 we did not find any  
 amino acid exchanges.  
 Clone distribution: http://www.rzpd.de/products/orfclones/.

# FEATURES

source  
 1..537  
 /location/Qualifiers  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="RZPD0834811D"  
 /clone\_lib="Human Full ORF Clones Gateway(TM) - RZPD"  
 /lab\_host="DH108"  
 /notes="Vector: pDONR201, Site\_1: attP1; Site\_2: attP2"  
 gene  
 1..537  
 /gene="TCF21"  
 CDS  
 1..537  
 /gene="TCF21"  
 /codon\_start=1  
 /protein\_id="CAG29289.1"  
 /db\_xref="GI:47496533"  
 /translation="MSGTSLSDVLDQVEMLCECDGLKMDNSKKEFVTNNESTERSSNC  
 ENGSPKQGGGLGKKRKATPKSPISGVQEGKQVORNAARERARMRVLSKAPSL  
 KTLTPWPPDTKLSDTLRLASSYIAHLRLANDKYENGYIHPVNLTFWFMVAGKP  
 ESDLKEVVFASRLCGTAS"

## ORIGIN

Alignment Scores:  
 Pred. No.: 1.65e-76 Length: 537  
 Score: 917.00 Matches: 179  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 9 Gaps: 0

US-09-701-674A-23 (1-179) x CR450293 (1-537)

QY 1 MetSerThrGlySerLeuSerAspValGluAspLeuGlnGluValGluMetLeuGluCys 20  
 DB 1 AYTCCACCGGCTCCCTCAGCATGTGGAGGACCTTCAAGAGTGGAGATGTTGAATGT 60

QY 21 AspGlyLeuLysMetAspSerAnLysGluPheValThrSerAsnGluSerThrGluGlu 40  
 DB 61 GACGGGTGAAATGATTCGAACAGGAATTTGTGACTTCCACGAGAGCACCAGGAG 120

QY 41 SerSerAsnCysGluAsnGlySerProGlnLysGlyArgGlyGlyLeuGlyLysArg 60  
 DB 121 AGTCTCAATCTGCAGAAATGGTCTCCCAAGAGGCGCGCGGCTCGGCAAGAGGAGG 180

QY 61 LysAlaProThrLysLysSerProLeuSerGlyValSerGlnGluGlyGlnValGln 80  
 DB 181 AAGGCGCCCAACCAAGAAGACCCCTCAGCGGGGTGACGAGGAGGAGGAGGTCACG 240

QY 81 ArgAsnAlaAsnAlaArgGluArgAlaArgMetArgValLeuSerLysAlaPheSer 100  
 DB 241 CGCAACGCGCCCAACGCGAGAGCGGGCGCCGATGCGAGTGTGAGCAAGCGCTTCTCC 300

QY 101 ArgLeuLysThrThrLeuProTyrValProProAspThrLysLeuSerLysLeuAspThr 120  
 DB 301 AGACTCAAGACCAACCTGCGTGGTGCCTCCACCAAGCTCTCCAAGCTGGACACG 360

QY 121 LeuArgLeuAlaSerSerTyrIleAlaHisLeuArgGlnIleLeuAlaAsnAspLysTyr 140  
 DB 361 CTCAGGCTGGGCTCAGCTACATCGCCCATCTGAGGAGATCTCTGGCTACGCAATAC 420

QY 141 GluAsnGlyTyrIleHisProValAsnLeuThrTrpProPheMetValAlaGlyLysPro 160  
 DB 421 GAGAACGGGTACATTCACCCGCTCAACCTGACGTGGCGCTTTATGTGGCGGGAACCC 480

QY 161 GluSerAspLeuLysGluValValThrAlaSerArgLeuCysGlyThrThrAlaSer 179  
 DB 481 GAGAGTGACCTGAAAGAAGTGGTGACCGGAGCGCTTATGTGGAACCAACCGCGTCC 537

RESULT 2  
 BD211771  
 LOCUS BD211771 1272 bp DNA linear PAT 17-JUL-2003  
 DEFINITION Proteins regulating gene expression.  
 ACCESSION BD211771  
 VERSION BD211771.1 GI:33021541  
 KEYWORDS JP 2002517246-A/23.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 1272)  
 AUTHORS Lal,P., Yue,H., Tang,Y.T., Hillman,J.L., Bandman,O., Corley,N.C.,  
 Guegler,K.J., Gorgone,G.A., Baughn,M.R., Patterson,C. and Lu,D.A.M.  
 TITLE Proteins regulating gene expression  
 JOURNAL Patent: JP 2002517246-A 23 18-JUN-2002;  
 COMMENT INCYTE PHARMACEUTICALS INC  
 OS Homo sapiens (human)  
 FN JP 2002517246-A/23  
 PD 18-JUN-2002  
 PF 11-JUN-1999 JP 200053586  
 PR 12-JUN-1998 US 60/089029, 29-JUL-1998 US 60/094575 PR  
 14-OCT-1998 US 60/104624  
 PI PREETI LAL, HENRY YUE, Y TOM TANG, JENNIFER L HILLMAN, OLGA PI  
 BANDMAN,  
 PI NEIL C CORLEY, KARL J GUEGLER, GINA A GORGONE, MARIAH R BAUGHN,  
 PI CHANDRA PATTERSON, DYUNG AINA M LU  
 PC C12N15/00, A61K38/00, A61K39/395, A61K39/395, A61K45/00, A61P15/00,  
 PC A61P25/00,  
 PC A61P35/00, A61P37/02, A61P43/00, C07K14/47, C07K16/18, C12N15/00,  
 PC A61K37/02  
 CC Incyte Clone No.: 3340296  
 FH Key location/Qualifiers  
 FT source 1..1272  
 FT /organism="Homo sapiens (human)".

## FEATURES

source  
 1..1272  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"

## ORIGIN

Alignment Scores:  
 Pred. No.: 4.86e-76 Length: 1272  
 Score: 917.00 Matches: 179  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 6 Gaps: 0

US-09-701-674A-23 (1-179) x BD211771 (1-1272)

QY 1 MetSerThrGlySerLeuSerAspValGluAspLeuGlnGluValGluMetLeuGluCys 20  
 DB 268 ATGTCCACCGGCTCCCTCAGCATGTGGAGGACCTTCAAGAGTGGAGATGTTGAATGT 327

QY 21 AspGlyLeuLysMetAspSerAnLysGluPheValThrSerAsnGluSerThrGluGlu 40

Db	328	GACGGTTCGAAATGATTCGACAGAGATTTGTGACTTCAACGAGGACCGAGGAG	387
QY	41	SerSerAsnCysGluAsnGlySerProGlnLysGlyArgGlyGlyLeuGlyLysArgArg	60
Db	388	AGCTCCAACTGGGAGAAATGGTCTCCCAAGAAAGGCGCGCGCTGGCAAGAGGAGG	447
QY	61	LysAlaProThrLysLysSerProLeuSerGlyValSerGlnGluGlyLysGlnValGln	80
Db	448	AGGCGCCCAACCAAGAGAGCCCTCAGCGGGGTTCAGCCAGAGGGAGACGAGTCCAG	507
QY	81	ArgAsnAlaAlaAsnAlaArgGluArgAlaArgMetArgValLeuSerLysAlaPheSer	100
Db	508	CGCAACCGCCCAACGCGGAGAGCGGGCCCGCATCGAGTCTCAGCAAGGCTTCTCC	567
QY	101	ArgLeuLysThrThrLeuProTrpValProProAspThrLysLeuSerLysLeuAspThr	120
Db	568	AGACTCAAGACCAACCTCGCCCTGGGTGGCCCGCCCGACCAACGCTTCAAGTGTGACACG	627
QY	121	LeuArgLeuAlaSerSerTyrrileAlaHisLeuArgGlnileLeuAlaAsnAspLysTyr	140
Db	628	CTCAGGCTGGCTCCAGCTACATCGCCCACTTGAGCGAGATCTGGCTAACGACAAATAC	687
QY	141	GluAsnGlyTyrrileHisProValLeuLeuThrTrpProPheMetValAlaGlyLysPro	160
Db	688	GAGAACGGGTACATTACCCCGGTCAACCTGACGTGGCCCTTTATGTGGCCCGGAAACCC	747
QY	161	GluSerAspLeuLysGluValValThrAlaSerArgLeuCysGlyThrThrAlaSer	179
Db	748	GAGAGTGACTGAAAGAGTGTGACCGGAGCGCGCTTATGTGACCAACCGCGTCC	804
RESULT 3			
BC025697			
LOCUS			
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
PUBLISHED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			

NIH-MGC project URL: http://mgc.nci.nih.gov  
Contact: MGC help desk  
Email: mgc@nci.nih.gov  
Tissue Procurement: Life Technologies, Inc.  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILML)  
DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC),  
Gaithersburg, Maryland;  
Web site: http://www.nisc.nih.gov/  
Contact: nisc\_mgc@nci.nih.gov  
Antker, N., Ayele, K., Beckstrom-Sternberg, S.M., Benjamin, B.,  
Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S.,  
Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghighi, P.,  
Hansen, N., Ho, S.-L., Karlins, E., Kwong, P., Laric, P., Legaspi, R.,  
Maduro, O.L., Masello, C., Maskeri, B., Mastrian, S.D., McCloskey, J.C.,  
McDowell, J., Pearson, R., Stantropop, S., Thomas, P.J., Touchman, J.W.,  
Turgeon, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L.,  
Young, A., Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found  
through the I.M.A.G.E. Consortium/ILML at: http://image.llnl.gov  
Series: IRAK Plate: 49 Row: k Column: 7  
This clone was selected for full length sequencing because it  
passed the following selection criteria: matched mRNA gi: 4507394.  
Location/Qualifiers  
1. 3231  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="MGC:34534 IMAGE:5221575"  
/tissue\_type="Pancreas, Spleen, adult pooled"  
/clone\_lib="NIH\_MGC\_120"  
/lab\_host="DH10B"  
/note="Vector: pCMV-SPORT6"  
1. 3231  
/gene="TCF21"  
/note="synonym: POD1"  
/db\_xref="LocusID:6943"  
/db\_xref="MIM:603306"  
260.799  
/gene="TCF21"  
/codon\_start=1  
/product="transcription factor 21"  
/protein\_id="AAH25697.1"  
/db\_xref="GI:19344016"  
/db\_xref="LocusID:6943"  
/db\_xref="MIM:603306"  
/translation="MSTGSLSDVEDLQVEMLCDGLMDSNKEFVTSNESTESSNC  
ENGSPKRGGLGKRRKAPTKSPKSGVQEGVQVORNAANARARMVLKAFSL  
KTLIPWPPTKLSKLDLTLASSYIAHLQILANDKVNGYIHPVNLTPFMVAGKP  
ESDLKEVVTASRLCGTTAS"

ORIGIN  
Alignment Scores:  
Pred. No.: 1.56e-75 Length: 3231  
Score: 917.00 Matches: 179  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 9 Gaps: 0  
US-09-701-674A-23 (1-179) x BC025697 (1-3231)  
QY 1 MetSerThrGlySerLeuSerAspValGluAspLeuGlnGluValGluMetLeuGluCys 20  
Db 260 ATGTCACCGCGCTCCCTCAGCGATGTGAGGACCTTCAAGAGGTGGAGATGTTGGAATGT 319  
QY 21 AspGlyLeuLysMetAspSerAsnLysGluPheValThrSerAsnGluSerThrGluGlu 40  
Db 320 GACGGTTCGAAATGATTCGACAGAGATTTGTGACTTCAACGAGGACCGAGGAG 379  
QY 41 SerSerAsnCysGluAsnGlySerProGlnLysGlyArgGlyGlyLeuGlyLysArgArg 60

```

Db      380 AGTCCAACTGCAGAAATGGTCTCCCAAGAGGCGCGCGCGCTGGCGCAAGAGGAGG 439
Qy      61 LysAlaProThrLysLysSerProLeuSerGlyValSerGlnGluGlyValGln 80
Db      440 AAGGCGCCCAACCAAGAGAGCCCTTGAGCGGGTTCAGCAGAGGAGGAGGAGGAGGTCAG 499
Qy      81 ArgAsnAlaAlaAsnAlaArgGluAlaArgMetArgValLeuSerLysAlaPheSer 100
Db      500 CGCAACGCCGCCCAACGCGAGAGCGGCGCGCATCGAGTGTGAGCAGAGGCTTCTCC 559
Qy      101 ArgLeuLysThrThrLeuProTyrValProProAspThrLysLysLysLysLysThr 120
Db      560 AGACTCAAGACCACTCCCTGCGGTCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 619
Qy      121 LeuArgLeuAlaSerGlyTrilleAlaHisLeuArgGlnIleLeuAlaAsnAspLysTyr 140
Db      620 CTCAGGCTGGCGTCCAGTACATCGCCACTTCGAGCGAGATCTGCTTAACGACAAATAC 679
Qy      141 GluAsnGlyTyrIleHisProValAsnLeuThrTrpProPheMetValAlaGlyLysPro 160
Db      680 GAGACGGGTACATTCAACCGGTCACCTGAGCGTGGCGCTTTATGGTGGCGCGGAAACCC 739
Qy      161 GluSerAspLeuLysGluValValThrAlaSerArgLeuCysGlyThrThrAlaSer 179
Db      740 GAGAGTGACCTGAAGAGTGGTGACCGCGCGCGCTTATGTGGAACCAACCGCGTCC 796

RESULT 4
AF035718 1254 bp mRNA linear PRI 05-OCT-1998
LOCUS     Homo sapiens mesoderm-specific basic-helix-loop-helix protein
DEFINITION
(POD1) mRNA, complete cds.
ACCESSION AF035718
VERSION    98175875
KEYWORDS   AF035718.1 GI:2745886
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1254)
AUTHORS   Quaggin,S.E., Vanden Heuvel,G.B. and Igarashi,P.
TITLE     Pod-1, a mesoderm-specific basic-helix-loop-helix protein expressed
in mesenchymal and glomerular epithelial cells in the developing
kidney
JOURNAL   Mech. Dev. 71 (1-2), 37-48 (1998)
MEDLINE   98175875
PUBMED    9507058
REFERENCE 2 (bases 1 to 1254)
AUTHORS   Quaggin,S.E., Vanden Heuvel,G.B. and Igarashi,P.
TITLE     Direct Submission
JOURNAL   Submitted (24-NOV-1997) Internal Medicine, Yale University, 333
Cedar Street, New Haven, CT 06520-8029, USA

FEATURES             source
1..1254
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/chromosomes="6"
gene               1..1254
/feature="POD1"
CDS                261..800
/feature="POD1"
/feature="POD1"
/feature="Pod-1"
/codon_start=1
/product="mesoderm-specific basic-helix-loop-helix
protein"
/protein_id="AAC62514.1"
/db_xref="GI:2745887"
/translation="WSTGSLSDVLEQVEMLECDGLKMDSNKEFVTSNESTERSNC
ENGSPQKGGGKRRAPTKSVSGVSGQKQVORNAANARARARVLSKAFSRU
KTLFWPPDTKLKLDTLRLASSYVIAHLRLANDKRYGIIHPVNLTFWFWAGKP
ESDLKEVTVASLRCGTTAS"
misc_feature       492..653

```

```

/feature="POD1"
/feature="encodes basic-helix-loop-helix domain"

ORIGIN
Alignment Scores:          9.08e-76          Length:          1254
Pred. No.:                914.00           Matches:          178
Score:                    100.00%          Conservative:      1
Percent Similarity:       99.44%          Mismatches:       0
Best Local Similarity:    99.67%          Indels:           0
Query Match:              9          Gaps:             0
DB:                        9

US-09-701-674A-23 (1-179) x AF035718 (1-1254)

Qy      1 MetSerThrGlySerLeuSerAspValGluAspLeuGlnGluValGluMetLeuGluCys 20
Db      261 ATGTCCACCGGCTCCCTCAGCGATGTGGAGCACTTCAAGAGGTGGAGATGTTGGAATGT 320
Qy      21 AspGlyLeuLysMetAspSerAsnLysGluPheValThrSerAsnGluSerThrGluGlu 40
Db      321 GACGGGTGAAATGGATTGAAACAGGAATTTGTGACTTCCAAACGAGAGCACCGAGGAG 380
Qy      41 SerSerAsnCysGluAsnGlySerProGlnLysGlyArgGlyGlyLeuGlyLysArgArg 60
Db      381 AGCTCCAACTGCGAGATGGTCTCTCCCAAGAGGCGCGCGCGCTGGGCAAGAGGAGG 440
Qy      61 LysAlaProThrLysLysSerProLeuSerGlyValSerGlnGluGlyLysGlnValGln 80
Db      441 AGGCGCGCCCAACGAGAGAGCGCCCTGAGCGGGTTCAGCCAGAGGAGGAGGAGGTCAG 500
Qy      81 ArgAsnAlaAlaAsnAlaArgGluAlaArgMetArgValLeuSerLysAlaPheSer 100
Db      501 CGCAACGCCGCCCAACGCGAGAGCGGCGCGCATGCGAGTGTGAGCAGAGGCTTCTCC 560
Qy      101 ArgLeuLysThrThrLeuProTyrValProProAspThrLysLysLysLysLysThr 120
Db      561 AGACTCAAGACCACTCCCTGCGGTCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 620
Qy      121 LeuArgLeuAlaSerGlyTrilleAlaHisLeuArgGlnIleLeuAlaAsnAspLysTyr 140
Db      621 CTCAGGCTGGCGTCCAGCTACATCGCCACTTCGAGCGAGATCTGCTTAACGACAAATAC 680
Qy      141 GluAsnGlyTyrIleHisProValAsnLeuThrTrpProPheMetValAlaGlyLysPro 160
Db      681 GAGAACGGGTACATTCAACCGGTCACCTGAGCGTGGCGCTTTATGGTGGCGGGAACCC 740
Qy      161 GluSerAspLeuLysGluValValThrAlaSerArgLeuCysGlyThrThrAlaSer 179
Db      741 GAGAGTGACCTGAAGAGTGGTGACCGGAGCGCGCTTATGTGGAACCAACCGCGTCC 797

RESULT 5
AF047419 1257 bp mRNA linear PRI 06-OCT-1998
LOCUS     Homo sapiens epicardin mRNA, complete cds.
DEFINITION
AF047419
ACCESSION AF047419.1 GI:2921852
VERSION    AF047419.1
KEYWORDS   Homo sapiens (human)
SOURCE     Homo sapiens
ORGANISM   Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1257)
AUTHORS   Robb,L., Mifsud,L., Hartley,L., Biben,C., Copeland,N.G.,
Gilbert,D.J., Jenkins,N.A. and Harvey,R.P.
TITLE     epicardin: A novel basic helix-loop-helix transcription factor gene
expressed in epicardium, branchial arch myoblasts, and mesenchyme
of developing lung, gut, kidney, and gonads
JOURNAL   Dev. Dyn. 213 (1), 105-113 (1998)
MEDLINE   98401941
PUBMED    9733105
REFERENCE 2 (bases 1 to 1257)
AUTHORS   Robb,L. and Mifsud,L.
TITLE     Direct Submission

```

Sept, bad date



JOURNAL Submitted (10-FEB-1998) Haematology and Cancer, Walter and Eliza Hall Institute of Medical Research, PO Royal Melbourne Hospital, Melbourne, VIC 3050, Australia

FEATURES source  
1 1257  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
CDS  
260..799  
/function="basic helix-loop-helix transcription factor"  
/note="expressed in embryonic epicardium and mesenchyme-derived tissues of lung, gut, kidney, and gonad"  
/codon\_start=1  
/product="epicardin"  
/protein\_id="AAC62532.1"  
/db\_xref="GI:2921853"  
/translation="MSTGSLSDVEDLOEVMLECDGLKMDSNKPFVTSNESTERSSNC  
ENGSPQKGGGLGKPKAPTKPSPLSGVSOEGKQVORNTANARERARVLSKAPSRLL  
KTLPPWPDPTKLSKLDLTLRLASSYIAHRLQILANDKYENGYIHPNLTWFPWVAGKP  
ESDLKEVVTASRLCGTTAS"

ORIGIN

Alignment Scores:  
Pred. No.: 1 13e-75 Length: 1257  
Score: 913.00 Matches: 178  
Percent Similarity: 99.44% Conservative: 0  
Best Local Similarity: 99.44% Mismatches: 1  
Query Match: 99.56% Indels: 0  
DB: 9 Gaps: 0

US-09-701-674A-23 (1-179) x AF047419 (1-1257)

QY 1 MetSerThrGlySerLeuSerAspValGluAspLeuGlnGluValGluMetLeuGluCys 20  
Db 260 ATGTCACCGGCTCCCTCAGCGATGTGGAGACCTTCAAGAGTGGAGATGTTGAATGT 319  
QY 21 AspGlyLeuLysMetAspSerAsnLysGluPheValThrSerAsnGluSerThrGluGlu 40  
Db 320 GACGGGTGMAATGATTCGACAGAGGATTTGTGACTTCCACGAGAGCCGAGGAG 379  
QY 41 SerSerAsnGlyLeuGlnGlySerProGlnLysGlyArgGlyLeuGlyLysArgArg 60  
Db 380 AGCTCCAACTGCGAGATGGCTCTCCAGAGAGGCGCGCGGCTGGGCAAGAGGAGG 439  
QY 61 LysAlaProThrLysLysSerProLeuSerGlyValSerGlnGluGlyLysGlnValGln 80  
Db 440 AAGCGCCCAACGAGAGAGCCCTCAGCGGGTCCAGCGAGGAGGAGGAGGAGGAGG 499  
QY 81 ArgAsnAlaAlaAsnAlaArgGluArgAlaArgMetArgValLeuSerLysAlaPheSer 100  
Db 500 CGCAACACCGCCACGCGAGAGCGGCGCGCATCGAGTGTGAGCAAGGCTTCTCC 559  
QY 101 ArgLeuLysThrThrLeuProThrValProProAspThrLysLeuSerLysLeuAspThr 120  
Db 560 AGACTCAAGACCACTCCCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 619  
QY 121 LeuArgLeuAlaSerSerTyrIleAlaHisLeuArgGlnIleLeuAlaAsnAspLysTyr 140  
Db 620 CTGAGCTGGCTCCAGTACATCGCCACTTGGAGCAGATCTGCTGAGCAACATAC 679  
QY 141 GluAsnGlyTyrIleHisProValAsnLeuThrTrpProPheMetValAlaGlyLysPro 160  
Db 680 GAGAACGGGTACATTCACCGGGTCAACTGACGTGGCGCTTTATGTTGGCGGGAACCC 739  
QY 161 GluSerAspLeuLysGluValValThrAlaSerArgLeuCysGlyThrThrAlaSer 179  
Db 740 GAGAGTGACCTGAAGAAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 796

RESULT 6  
CQ718038  
LOCUS  
DEFINITION Sequence 3972 from Patent WO02068579.  
1259 bp DNA linear PAT 03-FEB-2004

ACCESSION CQ718038  
VERSION CQ718038.1 GI:42278895  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1  
AUTHORS Venter, C.J., Adams, M.C., Li, P.W. and Myers, E.W.  
TITLE Kits, such as nucleic acid arrays, comprising a majority of  
human exons or transcripts, for detecting expression and other uses  
thereof  
JOURNAL Patent: WO 02068579-A 3972 06-SEP-2002;  
PE Corporation (NY) (US)  
FEATURES source  
1 1259  
/organism="Homo sapiens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"

ORIGIN

Alignment Scores:  
Pred. No.: 7 81e-75 Length: 1259  
Score: 904.00 Matches: 179  
Percent Similarity: 99.44% Conservative: 0  
Best Local Similarity: 99.44% Mismatches: 0  
Query Match: 98.58% Indels: 1  
DB: 6 Gaps: 0

US-09-701-674A-23 (1-179) x CQ718038 (1-1259)

QY 1 MetSerThrGlySerLeuSerAspValGluAspLeuGlnGluValGluMetLeuGluCys 20  
Db 264 ATGTCACCGGCTCCCTCAGCGATGTGGAGACCTTCAAGAGTGGAGATGTTGAATGT 323  
QY 21 AspGlyLeuLysMetAspSerAsnLysGluPheValThrSerAsnGluSerThrGluGlu 40  
Db 324 GACGGGTGMAATGATTCGACAGAGGATTTGTGACTTCCACGAGAGCCGAGGAG 383  
QY 41 SerSerAsnGlyLeuGlnGlySerProGlnLysGlyArgGlyLeuGlyLysArgArg 60  
Db 384 AGCTCCAACTGCGAGATGGTCTCCAGAGAGGCGCGGCTGGGCAAGAGGAGG 443  
QY 61 LysAlaProThrLysLysSerProLeuSerGlyValSerGlnGluGlyLysGlnValGln 80  
Db 444 AAGCGCCCAACGAGAGAGCCCTCAGCGGGTCCAGCGAGGAGGAGGAGGAGGAGG 503  
QY 81 ArgAsnAlaAlaAsnAlaArgGluArgAlaArgMetArgValLeuSerLysAlaPheSer 100  
Db 504 CGCAACCGCCCAACGCGAGAGCGGCGCGCATCGAGTGTGAGCAAGGCTTCTC 563  
QY 100 rArgLeuLysThrThrLeuProThrValProProAspThrLysLeuSerLysLeuAspThr 120  
Db 564 CAGACTCAAGACCACTCCCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 623  
QY 120 rLeuArgLeuAlaSerSerTyrIleAlaHisLeuArgGlnIleLeuAlaAsnAspLysTyr 140  
Db 624 GCTCAGGCTGGCTCCAGTACATCGCCACTTGGAGCAGATCTGCTGAGCAACATA 683  
QY 140 rGluAsnGlyTyrIleHisProValAsnLeuThrTrpProPheMetValAlaGlyLysP 160  
Db 684 CGAGAACGGGTACATTCACCGGGTCACTGACGTGGCGCTTTATGTTGGCGGGAAC 743  
QY 160 oGluSerAspLeuLysGluValValThrAlaSerArgLeuCysGlyThrThrAlaSer 179  
Db 744 CGAGAGTGACCTGAAGAAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 801

RESULT 7  
AF047418  
LOCUS  
DEFINITION Mus musculus epicardin mRNA, complete cds.  
ACCESSION AF047418  
VERSION AF047418.1 GI:2921850

## KEYWORDS

SOURCE Mus musculus (house mouse)

## ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 1202)

## REFERENCE

Robb, L., Mifsud, L., Hartley, L., Biben, C., Copeland, N.G.,  
Gilbert, D.J., Jenkins, N.A. and Harvey, R.P.

## AUTHORS

epicardin: A novel basic helix-loop-helix transcription factor gene  
expressed in epicardium, branchial arch myoblasts, and mesenchyme  
of developing lung, gut, kidney, and gonads  
Dev. Dyn. 213 (1), 105-113 (1998)

## JOURNAL

98401941

## MEDLINE

9733105

## PUBMED

2 (bases 1 to 1202)

## REFERENCE

Robb, L. and Mifsud, L.  
Submitted (10-FEB-1998) Haematology and Cancer, Walter and Eliza  
Hall Institute of Medical Research, PO Royal Melbourne Hospital,  
Melbourne, VIC 3050, Australia

## JOURNAL

Location/Qualifiers

## FEATURES

1..1202

/organism="Mus musculus"

/mol\_type="mRNA"

/strain="Balb/c"

/db\_xref="taxon:10090"

199..738

/function="basic helix-loop-helix transcription factor"

/notes="expressed in embryonic epicardium and

mesenchyme-derived tissues of lung, gut, kidney, and

gonad"

/codon\_start=1

/product="epicardin"

/protein\_id="AAC62531.1"

/db\_xref="GI:2921851"

/translation="MSTGLSDVEDLQVEMLDCSLKVDNSKFEFTGTSNESTEGSNC

ENGSPQKGGGLGKRRKAPTKKPSLGVSQEGQVORNAARERARVLSKAPSRLL

KTLPLWPPDTKLKSLDTLRASSYIAHLRQILANDKYENGYIHPVNLTPFWVAGKP

ENDLKEVVTANRLCGTTAS"

## ORIGIN

## Alignment Scores:

Pred. No.: 2,83e-73 Length: 1202  
Score: 887.00 Matches: 172  
Percent Similarity: 98.32% Conservative: 4  
Best Local Similarity: 96.09% Mismatches: 3  
Query Match: 96.73% Indels: 0  
DB: 10 Gaps: 0

US-09-701-674A-23 (1-179) x AF047418 (1-1202)

QY 1 MetSerThrGlySerLeuSerAspValGluAspLeuGlnValGluMetLeuGluCys 20  
DB 199 AUGTCCACCTGGCTCCCTCAGCGATGAGAGACCTTCAAGAGGTGGAGATCGTGCACGT 258  
QY 21 AspGlyLeuLysMetAspSerAsnLysGluPheValThrSerAsnGluSerThrGluGlu 40  
DB 259 GACTCCCTGAAAGTGGACTCCAAACAGGAGGTTTGGAACTTCCACGAGAGCACCAGGAG 318  
QY 41 SerSerAsnGluAsnGlySerProGlnLysGlyArgGlyGlyLeuGlyLysArgArg 60  
DB 319 GGCTCCAACTGGAGAACCGGCTCCACAGAGAGGTGCGGTGGCTAGGCAAGAGGAGG 378  
QY 61 LysAlaProThrLysLysSerProLeuSerGlyValSerGlnGluGlyLysGlnValGln 80  
DB 379 AAGGCCCCCACTAAGAAAACCGCTCAGCGGGTCCAGCCAGGAGGCGCAGAGTCCAG 438  
QY 81 ArgAsnAlaAlaAsnAlaArgGluAlaArgMetArgValLeuSerLysAlaPheSer 100  
DB 439 CGCAACGCGCAATGCTGCTGAGCGGCGCCCATGCGGGTCTGAGCAAGGCGCTTCTCC 498  
QY 101 ArgLeuLysThrThrLeuProTrpValProProAspThrLysLeuSerLysLeuAspThr 120

Db 499 AGGCTCAAGACCACCTGCTGGTGCCTCCCGGACACCAAGCTCTCCAAAGCTGGACACT 558  
QY 121 LeuArgLeuAlaSerSerTyrIleAlaHisLeuArgGlnIleLeuAlaAsnAspLysTyr 140  
DB 559 CTCAGGCTGGCTCAGCTACATCGCTTAAAGCAGATCTTGGCCAAACGAAAGTAC 618  
QY 141 GluAsnGlyTyrIleHisProValAsnLeuThrTrpProPheMetValAlaGlyLysPro 160  
DB 619 GAGAAAGGTTACATTACCCAGTCAACCTGACGTGGCCCTTTATGTGTGGCGGCAAAACA 678  
QY 161 GluSerAspLeuLysGluValThrAlaSerArgLeuCysGlyThrThrAlaSer 179  
DB 679 GAGAACTACCTGAAGGAAGTGGTACCGCCCAACCGCTTGTGTGGAACACTACAGCATCC 735

RESULT 8  
AF029753  
LOCUS 1217 bp mRNA linear ROD 25-FEB-2000  
DEFINITION Mus musculus basic helix-loop-helix factor Cor1 mRNA, complete cds.  
ACCESSION AF029753  
VERSION AF029753.1 GI:2613018  
KEYWORDS  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 1217)  
AUTHORS Hidai, H., Bardales, R., Goodwin, R., Quettermous, T. and  
Quettermous, E.E.  
TITLE Cloning of capsulin, a basic helix-loop-helix factor expressed in  
progenitor cells of the pericardium and the coronary arteries  
Mech. Dev. 73 (1), 33-43 (1998)  
JOURNAL 98213629  
MEDLINE 9545526  
PUBMED  
REFERENCE 2 (bases 1 to 1217)  
AUTHORS Quettermous, T., Quettermous, E.E., Hidai, H., Padel, B., Boutet, S.C.  
and Bardales, R.  
TITLE Direct Submission  
JOURNAL Submitted (10-OCT-1997) Department of Medicine, Division of  
Cardiology, Stanford University, Falk Bldg, 300 Pasteur Drive,  
Stanford, CA 94305, USA

FEATURES  
Location/Qualifiers  
1..1217  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
/tissue\_type="pericardium and coronary arteries"  
gene  
CDS  
1..1217  
/genes="Cor1"  
213..752  
/genes="Cor1"  
/function="transcription factor"  
/codon\_start=1  
/product="basic helix-loop-helix factor Cor1"  
/protein\_id="AAB84256.1"  
/db\_xref="GI:2613019"  
/translation="MSTGLSDVEDLQVEMLDCSLKVDNSKFEFTGTSNESTEGSNC  
ENGSPQKGGGLGKRRKAPTKKPSLGVSQEGQVORNAARERARVLSKAPSRLL  
KTLPLWPPDTKLKSLDTLRASSYIAHLRQILANDKYENGYIHPVNLTPFWVAGKP  
ENDLKEVVTANRLCGTTAS"

1026a

## ORIGIN

Alignment Scores:  
Pred. No.: 2,88e-73 Length: 1217  
Score: 887.00 Matches: 172  
Percent Similarity: 98.32% Conservative: 4  
Best Local Similarity: 96.09% Mismatches: 3  
Query Match: 96.73% Indels: 0  
DB: 10 Gaps: 0

US-09-701-674A-23 (1-179) x AF029753 (1-1217)

QY 1 MetSerThrGlySerLeuSerAspValGluAspLeuGlnValGluMetLeuGluCys 20





```
/clone_lib="NCI_CGAP_Co24"
/lab_host="DH108"
/notes="vector: PCMV-SPORT6"
1..1246
/gene="Tcf21"
/notes="synonyms: epc, Pod1; capsulin, epicardin, Pod-1"
/db_xref="LocusID:21412"
/db_xref="MGI:1202715"
233..772
/gene="Tcf21"
/codon_start=1
/product="transcription factor 21"
/protein_id="AAH53525.1"
/db_xref="GI:31566110"
/db_xref="LocusID:21412"
/db_xref="MGI:1202715"
/translation="MSTGSLSDVEDLOEVLDCSLKVDNKEFGTSNESTEGSNC
ENGSPQGRGLGKRRKAPTKKPSGVQSGKQVQRNANARERAMRVLSKAFSRL
KTLTPWPVPTKLSKLTDLRLASSYIAHLRQILANDKYENGYIHPVNLTPFWMVAGKP
ENDLKEVVTANRLCGTTAS"
ORIGIN
source
1..1267
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
264..803
/function="BHLH transcription factor"
/codon_start=1
/product="capsulin"
/protein_id="AAC23537.1"
/db_xref="GI:2695695"
/translation="MSTGSLSDVEDLOEVLDCSLKVDNKEFGTSNESTEGSNC
ENGSPQGRGLGKRRKAPTKKPSGVQSGKQVQRNANARERAMRVLSKAFSRL
KTLTPWPVPTKLSKLTDLRLASSYIAHLRQILANDKYENGYIHPVNLTPFWMVAGKP
ENDLKEVVTANRLCGTTAS"
Alignment Scores:
Pred. No.: 3,03e-73 Length: 1267
Score: 887.00 Matches: 172
Percent Similarity: 98.32% Conservative: 4
Best Local Similarity: 96.09% Mismatches: 3
Query Match: 96.73% Indels: 0
DB: 10 Gaps: 0
US-09-701-674A-23 (1-179) x BC053525 (1-1246)
QY 1 MetSerThrGlySerLeuSerAspValGluAspLeuGlnValGluMetLeuGluCys 20
DB 233 ATGTCCACTGGCTCCCTCAGCGATGTAGAACCTTCAAGAGGTGGAGATGCTGGACTGT 292
QY 21 AspGlyLeuLysMetAspSerAsnLysGluPheValThrSerAsnGluSerThrGluGlu 40
DB 293 GACTCCCTGAAAGTGAAGTCCCAAGAGGATTTGGAACTTCCCAAGAGGACCGAGGAG 352
QY 41 SerSerAsnCysGluAsnGlySerProGlnLysGlyValGlyClyLeuGlyLysArgArg 60
DB 353 GGCTCCCACTGCCAGAACGGGTCTCCACAGAGGGTCCGGTGGCTTAGCGAAGAGGAGG 412
QY 61 LysAlaProThrLysLysSerProLeuSerGlyValSerGlnGlnGlyLysGlnValGln 80
DB 413 AAGCGGCCCACTAAGAAAAGCCCGCTCAGCGGGGTCCAGCCAGGAGGCAAGCAGGTCCAG 472
QY 81 ArgAsnAlaAlaAsnAlaArgGluAlaArgMetArgValLeuSerLysAlaPheSer 100
DB 473 CGCAACGGGCCCAATGCTGTGAGCGGGCCCGCATCGGGTGTGTGAGCAAGCCCTTCTCC 532
QY 101 ArgLeuLysThrThrLeuProTIPValProProAspThrLysLeuSerLysLeuAspThr 120
DB 533 AGGCTCAAGACCACTCCCTGGGTGGCCCGGACCAAGCTCTCCAAAGCTGGACACT 592
QY 121 LeuArgLeuAlaSerSerTyrIleAlaHisLeuArgGlnIleLeuAlaAsnAspLysTyr 140
DB 593 CTCAGGTGGCGGCCAGCTACATCGCTCACTTAAGGCAGATCTCTGGCCCAACGACAGTAC 652
QY 141 GluAsnGlyTyrIleHisProValAsnLeuThrTIPProPheMetValAlaGlyLysPro 160
DB 653 GAGAACGGTTACATTCAACCAAGTCAACCTGACGTGGCCCTTATGGTGGCCGCAACCA 712
QY 161 GluSerAspLeuLysGluValValThrAlaSerArgLeuCysGlyThrThrAlaSer 179
DB 713 GAGAATGACCTGAAGAAAGTGTGACCGCAACCGCTTGTGTGGAACACTACAGCATCC 769
RESULT 12
AF036945 1267 bp mRNA linear ROD 15-JUN-1998
LOCUS Mus musculus capsulin mRNA, complete cds.
DEFINITION AF036945
ACCESSION
```

VERSION AF036945.1 GI:2695694  
KEYWORDS Mus musculus (house mouse)  
SOURCE Mus musculus  
ORGANISM Mus musculus  
REFERENCE 1 (bases 1 to 1267)  
AUTHORS Lu, J., Richardson, J.A. and Olson, E.N.  
TITLE Capsulin: a novel BHLH transcription factor expressed in epicardial progenitors and mesenchyme of visceral organs  
JOURNAL Mech. Dev. 73 (1), 23-32 (1998)  
MEDLINE 98213628  
PUBMED 9545521  
REFERENCE 2 (bases 1 to 1267)  
AUTHORS Lu, J. and Olson, E.N.  
TITLE Direct Submission  
JOURNAL Submitted (04-DEC-1997) Molecular Biology and Oncology, UT Southwestern Medical Center, 6000 Harry Hines Blvd., NAB.510, Dallas, TX 75235-9148, USA  
FEATURES  
Location/Qualifiers  
1..1267  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
264..803  
/function="BHLH transcription factor"  
/codon\_start=1  
/product="capsulin"  
/protein\_id="AAC23537.1"  
/db\_xref="GI:2695695"  
/translation="MSTGSLSDVEDLOEVLDCSLKVDNKEFGTSNESTEGSNC  
ENGSPQGRGLGKRRKAPTKKPSGVQSGKQVQRNANARERAMRVLSKAFSRL  
KTLTPWPVPTKLSKLTDLRLASSYIAHLRQILANDKYENGYIHPVNLTPFWMVAGKP  
ENDLKEVVTANRLCGTTAS"

Alignment Scores:  
Pred. No.: 3,03e-73 Length: 1267  
Score: 887.00 Matches: 172  
Percent Similarity: 98.32% Conservative: 4  
Best Local Similarity: 96.09% Mismatches: 3  
Query Match: 96.73% Indels: 0  
DB: 10 Gaps: 0

US-09-701-674A-23 (1-179) x AF036945 (1-1267)

QY 1 MetSerThrGlySerLeuSerAspValGluAspLeuGlnValGluMetLeuGluCys 20  
DB 264 ATGTCCACTGGCTCCCTCAGCGATGTAGAACCTTCAAGAGGTGGAGATGCTGGACTGT 323  
QY 21 AspGlyLeuLysMetAspSerAsnLysGluPheValThrSerAsnGluSerThrGluGlu 40  
DB 324 GACTCCCTGAAAGTGAAGTCCCAAGAGGATTTGGAACTTCCCAAGAGGACCGAGGAG 383  
QY 41 SerSerAsnCysGluAsnGlySerProGlnLysGlyValGlyClyLeuGlyLysArgArg 60  
DB 384 GGCTCCCACTCGGAGAACGGGTCTCCACAGAGGGTCCGGTGGCTTAGCGCAAGAGGAG 443  
QY 61 LysAlaProThrLysLysSerProLeuSerGlyValSerGlnGlnGlyLysGlnValGln 80  
DB 444 AAGCGGCCCACTAAGAAAAGCCCGCTCAGCGGGGTCCAGCAGGAGGCAAGCAGGTCCAG 503  
QY 81 ArgAsnAlaAlaAsnAlaArgGluArgAlaArgMetArgValLeuSerLysAlaPheSer 100  
DB 594 CGCAACGGGCCCAATGCTGTGAGCGGGCCCGCATCGGGTGTGTGAGCAAGCCCTTCTCC 563  
QY 101 ArgLeuLysThrThrLeuProTIPValProProAspThrLysLeuSerLysLeuAspThr 120  
DB 564 AGGCTCAAGACCACTCCCTGGGTGGCCCGGACCAAGCTCTCCAAAGCTGGACACT 623  
QY 121 LeuArgLeuAlaSerSerTyrIleAlaHisLeuArgGlnIleLeuAlaAsnAspLysTyr 140  
DB 624 CTCAGGTGGCGGTCCAGCTACATCGCTCACTTAAGGCAGATCTCTGGCCCAACGACAGTAC 683

141 GluAsnGlyTyrIleHisProValAsnLeuThrTrpProPheMetValAlaGlyLysPro 160  
 684 GAGACGGTTTACATTACCCAGTCAACCTGACGTGGCCCTTTATGTCGCCGCAACCA 743  
 161 GluSerAspLeuLysGluValValThrAlaSerArgLeuCysGlyThrThrAlaSer 179  
 744 GAGAAATGACCTGAAGGAAGTGGTGACCGCAACCGCTTGTGTGGAACACTACACATCC 800  
 RESULT 13  
 Bx932800  
 LOCUS Bx932800 1164 bp mRNA linear VRT 30-MAR-2004  
 DEFINITION Gallus gallus finished cDNA, clone CHEST265h24.  
 ACCESSION Bx932800  
 VERSION Bx932800.2 GI:46017571  
 KEYWORDS  
 SOURCE Gallus gallus (chicken)  
 ORGANISM Gallus gallus  
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 Archosauria; Aves; Neognathae; Galliformes; Phasianidae;  
 Phasianinae; Gallus.  
 1 (bases 1 to 1164)  
 Boardman, P.E., Bonfield, J.K., Brown, W.R.A., Carder, C., Chalk, S.E.,  
 Croning, M.D.R., Davies, R.M., Francis, M.D., Grafton, D.V.,  
 Hubbard, S.J., Humphray, S.J., Hunt, P.J., Maddison, M., McLaren, S.R.,  
 Niblett, D., Overton, I.M., Rogers, J., Scott, C.E., Taylor, R.G.,  
 Tickie, C., and Wilson, S.A.  
 Direct Submission  
 Submitted (29-MAR-2004) Sanger Institute, Hinxton, Cambridgeshire,  
 CB10 1SA, UK. E-mail enquiries: chickst@ms.unist.ac.uk  
 On Apr 1, 2004 this sequence version replaced gi:41633328.  
 BSRG/Dundee/Nottingham/Sanger/Sheffield/UMIST Gallus gallus cDNA  
 sequencing project.  
 This sequence is from the  
 BSRG/Dundee/Nottingham/Sanger/Sheffield/UMIST cDNA collection,  
 from a library constructed by Elizabeth Bosch. cDNA was prepared  
 from RNA extracted from ovary, normalised, and poly A-trimmed.  
 EcoRI-NotI cut cDNA was then ligated into the vector. Vector:  
 pBluescript II KS(+); Site\_1: EcoRI; Site\_2: NotI Host: Escherichia  
 coli DH10B.  
 FEATURES  
 source  
 1. 1164  
 /organism="Gallus gallus"  
 /mol\_type="mRNA"  
 /strain="Layer"  
 /db\_xref="taxon:9031"  
 /clone="CHEST265h24"  
 /clone\_lib="CSEORBN19"  
 /dev\_stage="adult"  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 8 136-69 Length: 1164  
 Score: 839.00 Matches: 164  
 Percent Similarity: 93.85% Conservative: 4  
 Best Local Similarity: 91.62% Mismatches: 11  
 Query Match: 91.49% Indels: 0  
 DB: 5 Gaps: 0  
 US-09-701-674A-23 (1-179) x Bx932800 (1-1164)  
 1 MetSerThrGlySerLeuSerAspValGluAspLeuGlnGluValMetLeuGluCys 20  
 167 ATGTCACATGGGTCCTCGATGATGGAAGATCTCTCAAGAGGTGGAATGTCGAGTGC 226  
 21 AspGlyLeuLysMetAspSerAsnLysGluPheValThrSerAsnGluSerThrGluGlu 40  
 227 GATGGCTGAAATGATGATACACAGAGTTCGGGGCTCCAAAGAGACACAGAGAG 286  
 41 SerSerAsnCysGluAsnGlySerProGlnLysGlyArgGlyGlyLeuGlyLysArg 60  
 287 GGATCCAAATGGCGAGAATGGCTCCCTCAGAAAGGGAGAGGGCGCTCGGGAAGAGGAA 346

61 LysAlaProThrIlyLysLysSerProLeuSerGlyValSerGlnGluGlyLysGlnValGln 80  
 347 AAGGCTCCCCCAAGAGAGCCCTTTAAATGGTGTGAGCCAGGAAGAAAGCAGGTCAG 406  
 81 ArgSerAlaAlaSerAlaArgGluArgAlaArgMetArgValLeuSerLysAlaPheSer 100  
 407 AGAAACGCTGCCAACGCAAGGAGAGGGCAAGATAGGGTCTTTAGCAAGCTTCTCC 466  
 101 ArgLeuLysThrThrLeuProTyrValProAspThrLysLeuSerLysLeuAspThr 120  
 467 AGGCTTAAGACCACTGCTGCTGGTGCCTCCAGACACCAAGCTTCCAACTGGACACC 526  
 121 LeuArgLeuAlaSerSerTyrIleAlaHisLeuArgGlnIleLeuAlaSerLysTyr 140  
 527 TTGAGGCTGGCTCCAGCTACATCGCACCTGAGGCAAAATCTGCGCAACGCAAGTAC 586  
 141 GluAsnGlyTyrIleHisProValAsnLeuThrTrpProPheMetValAlaGlyLysPro 160  
 587 GAAATGGCTACATCCATCCAGTCAACCTGACCTGGCTTTATGTCGCGCAACACC 646  
 161 GluSerAspLeuLysGluValValThrAlaSerArgLeuCysGlyThrThrAlaSer 179  
 647 GAGAGTGACCTGAAGGAAGTGGTGAACACAAACCGCTGTGCGCGCCGCGGATCC 703  
 RESULT 14  
 Bx932800  
 LOCUS Bx932800 965 bp mRNA linear VRT 03-AUG-2004  
 DEFINITION Xenopus laevis trans transcription factor 21, mRNA (cDNA clone MGC:82899  
 IMAGE:5512805), complete cds.  
 ACCESSION Bx932800  
 VERSION Bx932800.1 GI:49256164  
 KEYWORDS  
 SOURCE MGC  
 ORGANISM Xenopus laevis (African clawed frog)  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;  
 Xenopodinae; Xenopus; Xenopus.  
 1 (bases 1 to 965)  
 Klein, S.L., Strausberg, R.L., Wagner, L., Pontius, J., Clifton, S.W.  
 and Richardson, P.  
 Genetic and genomic tools for Xenopus research: The NIH Xenopus  
 Initiative  
 Dev. Dyn. 225 (4), 384-391 (2002)  
 12454917  
 2 (bases 1 to 965)  
 Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,  
 Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,  
 Altschul, S.F., Zeeberg, B., Suetow, K.H., Schaefer, C.F., Bhat, N.K.,  
 Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F.,  
 Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,  
 Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,  
 Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S.,  
 Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J.,  
 Abramson, R.D., Mullahy, S.J., Bosak, S.A., McEwan, P.J.,  
 McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,  
 Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,  
 Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,  
 Fahey, J., Helton, E., Kettelman, M., Madan, A., Rodriguez, S.,  
 Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y.,  
 Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,  
 Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,  
 Butterfield, Y.S., Krzywicki, M.I., Skalska, U., Smaluk, D.E.,  
 Schnurch, A., Schein, J.E., Jones, S.J. and Marra, M.A.  
 Generation and initial analysis of more than 15,000 full-length  
 human and mouse cDNA sequences  
 Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)  
 12477932  
 3 (bases 1 to 965)  
 Klein, S. and Gerhard, D.S.  
 Direct Submission  
 Submitted (14-JUN-2004) National Institutes of Health, Xenopus Gene  
 Collection (XGC), National Institute of Child Health and Human  
 Development, 6100 Executive Boulevard, Room 4B01, Rockville, MD

Qy	61	LyAlaProThrIysIysSerProLeuSerGlyValserGlnGluGlyLysGlnValGln	80
		::: ::::	
Db	204	AAAGCCTCCAGCAAGAAAATGCCACTGGGCACCATCAACCAAGGAGGCAACAGGTCAG	263
Qy	81	ArgAsnAlaAlaAsnAlaArgGluArgAlaArgMetArgValLeuSerLysAlaPheSer	100
Db	264	AGGAATGCACCCAAATGCCAGGGAGAGCCAGATAGGGTGCTTAAGCAAGCCCTTTTCT	323
Qy	101	ArgLeuIysThrThrLeuProTyrValProAspThrLysLeuSerLysLeuAspThr	120
Db	324	AGGCTAAAGACCACTTTGCCCTGGTGCTCCAGACACCAAACTTTCCAAGTCGGACACT	383
Qy	121	LeuArgLeuAlaSerSerTyrIleAlaHisLeuArgGlnIleLeuAlaAsnAspLysTyr	140
Db	384	CTCGCGCTTCGACTCTAGTTACATAGTTCACCTCGCTCAAAATCTTAGCAATGACAAGTAT	443
Qy	141	GluAsnGlyTyrIleHisProValAsnLeuThrTyrProPheMetValAlaGlyLysPro	160
Db	444	GAANAATGGCTACATCCATCCAGTTAACTGACTTGGCCCTTTATGGTGGCTGGAAACCA	503
Qy	161	GluSerAspLeuLysGluValValThrAlaSerArgLeuCysGlyThrThrAlaSer	179
		::: ::::	
Db	504	GAACAACACCTCAAGAAGTGGTGAGCAGCTAGCGATTGTGTGTGTCCACCGGCATCT	560

RESULT 15					
AY660871	AY660871	1059 bp	mRNA	linear	VRT 20-JUL-2004
LOCUS	Xenopus laevis transcription factor 21 (Tcf2l)				
DEFINITION	Xenopus laevis transcription factor 21 (Tcf2l) mRNA, complete cds.				
ACCESSION	AY660871				
VERSION	AY660871.1 GI:50313157				
KEYWORDS	.				
SOURCE	Xenopus laevis (African clawed frog)				
ORGANISM	Xenopus laevis				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
	Ambiphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;				
	Xenopodinae; Xenopus; Xenopus.				
	1 (bases 1 to 1059)				
REFERENCE	Eid, S.R. and Brandli,A.W.				
AUTHORS	Essential function of the basic helix-loop-helix transcription				
TITLE	factor Tcf21 (capsulin/epicardin/Pod1) for Xenopus pronephric				
	kidney organogenesis				
JOURNAL	Unpublished				
REFERENCE	2 (bases 1 to 1059)				
AUTHORS	Eid, S.R. and Brandli,A.W.				
TITLE	Direct Submission				
JOURNAL	Submitted (21-JUN-2004) Institute of Pharmaceutical Sciences, Swiss Federal Institute of Technology (ETHZ), Winterthurerstrasse 190, Zurich CH-8057, Switzerland				
FEATURES	Location/Ovalifiers				

```

1..1059
location=Vd4a11r1c1s
/organism="Xenopus laevis"
/mol_type="mRNA"
/db_xref="taxon:8355"
/clone="Pod1-12b"
/tissue_type="kidney"
/dev_stage="adult"
1..1059
/gene="Tcf21"
134..673
/gene="Tcf21"
/notes="basic helix-loop-helix transcription factor;
capsulin; epicardin; podocyte-expressed 1; Pod1; Pod-1"
/codon_start=1
/product="transcription factor 21"
/protein_id="AA74527.1"
/db_xref="GI:50313358"
/translation="MTGTSQSDVEDFDQDMLECDGKLDPNKEFGISNDSNRSSTC
DNQSPKKGRTGSKGRKASKKPLGTINQEQVQQRNANARARMLVLSKAFSR
KTLVPVPPDTLKSILDTLKLASSYIAHLRQILANDKYENGYIHPVNLTFWPFMVAGFE
ENDLKEVYSTSRLCGPTAS"

```

Qy	1	MetSerThrGlySerLeuSerAspValGluAspLeuGlnGluValGluMetLeuGluCys	20
Db	24	ATGTCCACCGGTCTCTCAGTGATGTGGAGACTTCCAGACATCGGAATGTTGGAAATG	83
Qy	21	AspGlyLeuIysMetAspSerAsnLysGluPheValThrSerAsnGluSerThrGluGlu	40
Db	84	GATGGCATCAACATGGATCTCTAATAAGAAATTTGGCATATCCAACGACGACCAATGAGAG	143
Qy	41	SerSerAsnCysGluAsnGlySerProGlnLysGlyArgGlyGlyLeuGlyLysArg	60
Db	144	AGTTCACCTTCGCAACACGGTTCCACCAAAAAGCTGAGAGGACTTCGGGCAAGAGAGA	203

D<sub>b</sub> 144 AGTTCAACCTGCGACCAACGGTTCACCCAAAAGGTAGAGGGACTTCGGGCAAGAGGAGA 203



## Alignment Scores:

Pred. No.:	1.32e-65	Length:	1059
Score:	804.00	Matches:	154
Percent Similarity:	92.74%	Conservative:	12
Best Local Similarity:	86.03%	Mismatches:	13
Query Match:	87.68%	Indels:	0
DB:	5	Gaps:	0

US-09-701-674A-23 (1-179) x AY660871 (1-1059)

Qy	1	MetSerThrGlySerLeuSerAspValGluAspLeuGlnGluValGluMetLeuGluCys	20
Db	134	ATGTCACCGGTTCTCTAGTGTGGAGGACTTCCAAGACATGGAATGTGGAAATGT	193
Qy	21	AspGlyLeuLysMetAspSerAsnLysGluPheValThrSerAsnGluSerThrGluGlu	40
Db	194	GATGGCATCAAACTGGATCCTTAATAAGAAATTGGCATATCCAACGACAGCAATGAGGAG	253
Qy	41	SerSerAsnCysGluAsnGlySerProGlnLysGlyArgGlyLeuGlyLysArgArg	60
Db	254	AGTTCACCTGGCACAACCGTTGGCCCAAAAAGGTAGAGGACTTCGGGCCAAGAGGAGA	313
Qy	61	LysAlaProThrLysLysSerProLeuSerGlyValSerGlnGluGlyLysGlnValGln	80
Db	314	AAAGCCTCCAGCAAGAAAGTCCACTGGGCACCATCAACGAGGAGGCAAGCAGGTCCTCAG	373
Qy	81	ArgAsnAlaAlaAsnAlaArgGluArgAlaArgMetArgValLeuSerLysAlaPheSer	100
Db	374	AGGAATGCAGCAATCCAGGAGAGAGCCAGGATGAGGGTGCTAAGCAAAAGCCTTTTCT	433
Qy	101	ArgLeuLysThrThrLeuProTrpValProProAspThrLysLeuSerLysLeuAspThr	120
Db	434	AGSCTAAAGACCACTTTGGCTTGGGTGGCTCCAGACCCAACTTTCCAGCTGGACACT	493
Qy	121	LeuArgLeuAlaSerSerTyrlaHisLeuArgGlnIleLeuAlaAsnAspLysTyr	140
Db	494	CTGGGCTTGCATCTAGTTACATAGCTACCTGGCTCAATTTCTAGCCAATGACAAGTAT	553
Qy	141	GluAsnGlyTyrlaHisProValAsnLeuThrTrpProPheMetValaGlyLysPro	160
Db	554	GAAAATGGCTACATCCATCCAGTTAACCTTGACTTGGCCCTTTTATGGTGGCTGGAACCA	613
Qy	161	GluSerAspLeuLysGluValThrAlaSerArgLeuCysGlyThrThrAlaSer	179
Db	614	GAAACGACCTCAAGGAAGTGTGACCACTAGCCGATTGTGTGTCCTCCACGGCATCT	670

Search completed: December 16, 2004, 23:15:58

Job time : 3803 secs

XX  
 22  
 23  
 24  
 25  
 26  
 27  
 28  
 29  
 30  
 31  
 32  
 33  
 34  
 35  
 36  
 37  
 38  
 39  
 40  
 41  
 42  
 43  
 44  
 45  
 46  
 47  
 48  
 49  
 50  
 51  
 52  
 53  
 54  
 55  
 56  
 57  
 58  
 59  
 60  
 61  
 62  
 63  
 64  
 65  
 66  
 67  
 68  
 69  
 70  
 71  
 72  
 73  
 74  
 75  
 76  
 77  
 78  
 79  
 80  
 81  
 82  
 83  
 84  
 85  
 86  
 87  
 88  
 89  
 90  
 91  
 92  
 93  
 94  
 95  
 96  
 97  
 98  
 99  
 100  
 101  
 102  
 103  
 104  
 105  
 106  
 107  
 108  
 109  
 110  
 111  
 112  
 113  
 114  
 115  
 116  
 117  
 118  
 119  
 120  
 121  
 122  
 123  
 124  
 125  
 126  
 127  
 128  
 129  
 130  
 131  
 132  
 133  
 134  
 135  
 136  
 137  
 138  
 139  
 140  
 141  
 142  
 143  
 144  
 145  
 146  
 147  
 148  
 149  
 150  
 151  
 152  
 153  
 154  
 155  
 156  
 157  
 158  
 159  
 160  
 161  
 162  
 163  
 164  
 165  
 166  
 167  
 168  
 169  
 170  
 171  
 172  
 173  
 174  
 175  
 176  
 177  
 178  
 179  
 180  
 181  
 182  
 183  
 184  
 185  
 186  
 187  
 188  
 189  
 190  
 191  
 192  
 193  
 194  
 195  
 196  
 197  
 198  
 199  
 200  
 201  
 202  
 203  
 204  
 205  
 206  
 207  
 208  
 209  
 210  
 211  
 212  
 213  
 214  
 215  
 216  
 217  
 218  
 219  
 220  
 221  
 222  
 223  
 224  
 225  
 226  
 227  
 228  
 229  
 230  
 231  
 232  
 233  
 234  
 235  
 236  
 237  
 238  
 239  
 240  
 241  
 242  
 243  
 244  
 245  
 246  
 247  
 248  
 249  
 250  
 251  
 252  
 253  
 254  
 255  
 256  
 257  
 258  
 259  
 260  
 261  
 262  
 263  
 264  
 265  
 266  
 267  
 268  
 269  
 270  
 271  
 272  
 273  
 274  
 275  
 276  
 277  
 278  
 279  
 280  
 281  
 282  
 283  
 284  
 285  
 286  
 287  
 288  
 289  
 290  
 291  
 292  
 293  
 294  
 295  
 296  
 297  
 298  
 299  
 300  
 301  
 302  
 303  
 304  
 305  
 306  
 307  
 308  
 309  
 310  
 311  
 312  
 313  
 314  
 315  
 316  
 317  
 318  
 319  
 320  
 321  
 322  
 323  
 324  
 325  
 326  
 327  
 328  
 329  
 330  
 331  
 332  
 333  
 334  
 335  
 336  
 337  
 338  
 339  
 340  
 341  
 342  
 343  
 344  
 345  
 346  
 347  
 348  
 349  
 350  
 351  
 352  
 353  
 354  
 355  
 356  
 357  
 358  
 359  
 360  
 361  
 362  
 363  
 364  
 365  
 366  
 367  
 368  
 369  
 370  
 371  
 372  
 373  
 374  
 375  
 376  
 377  
 378  
 379  
 380  
 381  
 382  
 383  
 384  
 385  
 386  
 387  
 388  
 389  
 390  
 391  
 392  
 393  
 394  
 395  
 396  
 397  
 398  
 399  
 400  
 401  
 402  
 403  
 404  
 405  
 406  
 407  
 408  
 409  
 410  
 411  
 412  
 413  
 414  
 415  
 416  
 417  
 418  
 419  
 420  
 421  
 422  
 423  
 424  
 425  
 426  
 427  
 428  
 429  
 430  
 431  
 432  
 433  
 434  
 435  
 436  
 437  
 438  
 439  
 440  
 441  
 442  
 443  
 444  
 445  
 446  
 447  
 448  
 449  
 450  
 451  
 452  
 453  
 454  
 455  
 456  
 457  
 458  
 459  
 460  
 461  
 462  
 463  
 464  
 465  
 466  
 467  
 468  
 469  
 470  
 471  
 472  
 473  
 474  
 475  
 476  
 477  
 478  
 479  
 480  
 481  
 482  
 483  
 484  
 485  
 486  
 487  
 488  
 489  
 490  
 491  
 492  
 493  
 494  
 495  
 496  
 497  
 498  
 499  
 500  
 501  
 502  
 503  
 504  
 505  
 506  
 507  
 508  
 509  
 510  
 511  
 512  
 513  
 514  
 515  
 516  
 517  
 518  
 519  
 520  
 521  
 522  
 523  
 524  
 525  
 526  
 527  
 528  
 529  
 530  
 531  
 532  
 533  
 534  
 535  
 536  
 537  
 538  
 539  
 540  
 541  
 542

protein regulating gene expression PRGE-23 (see AY58630). The cDNA was initially isolated from spleen tissue cDNA library SPLNNOT10, and the full-length sequence assembled from overlapping clones from a number of libraries. PRGE-23 is expressed in reproductive, developmental and urologic tissues associated with cancer, inflammation and foetal diseases, disorders or conditions. It is characterized as an Mvz-type HLH protein. The invention provides PRGE polypeptides (see AY58608-38) and polynucleotides (see AAZ57839-69), expression vectors, host cells, antibodies, agonists and antagonists. It also provides methods for diagnosing, treating or preventing disorders associated with expression of PRGE. Polynucleotides are also used as sources of probes and primers for diagnosis and monitoring of disease, also for detecting related sequences and in gene mapping

Sequence 1272 BP; 311 A; 358 C; 318 G; 285 T; 0 U; 0 Other;

Query Match 100.0%; Score 1272; DB 3; Length 1272;  
 Best Local Similarity 100.0%; Pred. No. 5,1e-284;  
 Matches 1272; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

CC 1 TCTACGGCCACGACTCTGGAGTGGGAAACAGAGAGCGGTTCTCTGTCGCAAGTC 60  
 CC  
 CC 1 TCTACGGCCACGACTCTGGAGTGGGAAACAGAGAGCGGTTCTCTGTCGCAAGTC 60  
 CC  
 CC 61 CTCGGGGTTCCTTCTCAACACTCTGCAAGGGAAGGGTTGTGAGACCCACAGAGCC 120  
 CC  
 CC 61 CTCGGGGTTCCTTCTCAACACTCTGCAAGGGAAGGGTTGTGAGACCCACAGAGCC 120  
 CC  
 CC 121 CAATCCAGTCCACAGAGGTGCTGCGCCACACTCGGAGGCGCTTGGTTTCAGG 180  
 CC  
 CC 121 CAATCCAGTCCACAGAGGTGCTGCGCCACACTCGGAGGCGCTTGGTTTCAGG 180  
 CC  
 CC 181 GTCTCTCTCTCTCTCTCACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 240  
 CC  
 CC 181 GTCTCTCTCTCTCTCTCACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 240  
 CC  
 CC 241 CT 300  
 CC  
 CC 241 CT 300  
 CC  
 CC 301 CTTCAAGAGTGGAGATGTTGNAATGTCAGCGGTGAAATGATTCGAAACAGGAATT 360  
 CC  
 CC 301 CTTCAAGAGTGGAGATGTTGNAATGTCAGCGGTGAAATGATTCGAAACAGGAATT 360  
 CC  
 CC 361 GTGACTTCCAAACGAGAGCCAGGAGAGTCCAACTGCGAGATGGGTCTCTCCAGAG 420  
 CC  
 CC 361 GTGACTTCCAAACGAGAGCCAGGAGAGTCCAACTGCGAGATGGGTCTCTCCAGAG 420  
 CC  
 CC 421 GCGCGGGCGGCTCTGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480  
 CC  
 CC 421 GCGCGGGCGGCTCTGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480  
 CC  
 CC 481 GTGAGCAG 540  
 CC  
 CC 481 GTGAGCAG 540  
 CC  
 CC 541 ATGCGAGTGTGAGCAAGGCTTCTCCAGACTCAAGACCACTGCGGTGGTGGTGGT 600  
 CC  
 CC 541 ATGCGAGTGTGAGCAAGGCTTCTCCAGACTCAAGACCACTGCGGTGGTGGTGGT 600  
 CC  
 CC 601 GACACCAAGTCTCAAGTGGACAGCTCAGGCTGGGCTGAGCTAGATGCCACATG 660  
 CC  
 CC 601 GACACCAAGTCTCAAGTGGACAGCTCAGGCTGGGCTGAGCTAGATGCCACATG 660  
 CC  
 CC 661 AGGAGAGTCTGGCTAACGACAAATACGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 720  
 CC  
 CC 661 AGGAGAGTCTGGCTAACGACAAATACGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 720  
 CC  
 CC 721 TGCCCTTTATGTTGGCGGGAACCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 780  
 CC  
 CC 721 TGCCCTTTATGTTGGCGGGAACCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 780  
 CC  
 CC 781 CGCTTATGTGAACCAACCGCGTCTCTGACCTTGGAGGTGGAGTCTGGGAAAGCGCGTC 840  
 CC

781 CGCTTATGTGAACCAACCGCGTCTCTGACCTTGGAGTGGAGTCTGGAAAGCGCGTC 840  
 QY  
 841 CCAGGGGAGAGCGGCGCGCGGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900  
 DB  
 841 CCAGGGGAGAGCGGCGCGCGGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900  
 QY  
 901 CCCTCGCAATGCT 960  
 DB  
 901 CCCTCGCAATGCT 960  
 QY  
 961 ATTGCTTTCCAAACAGAGAGAGATCAATGTACTTACAAAGATTCCCATCTATTAACT 1020  
 DB  
 961 ATTGCTTTCCAAACAGAGAGAGATCAATGTACTTACAAAGATTCCCATCTATTAACT 1020  
 QY  
 1021 TATTAATCTTACCGTGAATGACTCTGCAAGCCTTCTGCTGCTCAAGTGAATATGTAAT 1080  
 DB  
 1021 TATTAATCTTACCGTGAATGACTCTGCAAGCCTTCTGCTGCTCAAGTGAATATGTAAT 1080  
 QY  
 1081 ATAAATATATAAATAGATAGAGCCTATCAATGTATCTTTGTACAAATATGTTGTAAT 1140  
 DB  
 1081 ATAAATATATAAATAGATAGAGCCTATCAATGTATCTTTGTACAAATATGTTGTAAT 1140  
 QY  
 1141 GTAGATCATAGGATAGTGTGACTTTGACAGTCACTTTATAAGTAATCACTTAAAGATA 1200  
 DB  
 1141 GTAGATCATAGGATAGTGTGACTTTGACAGTCACTTTATAAGTAATCACTTAAAGATA 1200  
 QY  
 1201 TATATTTTTTCAACAAGTTTGTCTACTTTGCAAAATATCTTTTATATATCTAA 1260  
 DB  
 1201 TATATTTTTTCAACAAGTTTGTCTACTTTGCAAAATATCTTTTATATATCTAA 1260  
 QY  
 1261 AAAAAAAAAA 1272  
 DB  
 1261 AAAAAAAAAA 1272

RESULT 2  
 ABL65519  
 ID ABL65519 standard; DNA; 697 BP.  
 XX  
 AC ABL65519;  
 XX  
 DT 15-MAY-2002 (first entry)  
 XX  
 DE Lung cancer related gene sequence SEQ ID NO:3856.  
 XX  
 KW Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;  
 KW stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;  
 KW cytoskeletal; Gene therapy; antineoplastic; Wilm's tumour; adenocarcinoma;  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200194629-A2.  
 XX  
 PD 13-DEC-2001.  
 XX  
 PP 30-MAY-2001; 2001WO-US010838.  
 XX  
 PR 05-JUN-2000; 2000US-0209473P.  
 PR 05-JUN-2000; 2000US-0209531P.  
 PR 18-SEP-2000; 2000US-0233133P.  
 PR 18-SEP-2000; 2000US-023317P.  
 PR 20-SEP-2000; 2000US-0234009P.  
 PR 20-SEP-2000; 2000US-0234034P.  
 PR 20-SEP-2000; 2000US-0234052P.  
 PR 22-SEP-2000; 2000US-0234509P.  
 PR 22-SEP-2000; 2000US-0234567P.  
 PR 25-SEP-2000; 2000US-0234923P.  
 PR 25-SEP-2000; 2000US-0234924P.  
 PR 25-SEP-2000; 2000US-0235077P.  
 PR 25-SEP-2000; 2000US-0235082P.  
 PR 25-SEP-2000; 2000US-0235134P.



PR 25-SEP-2000; 2000US-0235082P.  
PR 25-SEP-2000; 2000US-0235134P.  
PR 25-SEP-2000; 2000US-0235280P.  
PR 26-SEP-2000; 2000US-0235637P.  
PR 26-SEP-2000; 2000US-0235638P.  
PR 27-SEP-2000; 2000US-0235711P.  
PR 27-SEP-2000; 2000US-0235720P.  
PR 27-SEP-2000; 2000US-0235840P.  
PR 27-SEP-2000; 2000US-0235863P.  
PR 28-SEP-2000; 2000US-0236028P.  
PR 28-SEP-2000; 2000US-0236032P.  
PR 28-SEP-2000; 2000US-0236033P.  
PR 28-SEP-2000; 2000US-0236034P.  
PR 28-SEP-2000; 2000US-0236109P.  
PR 28-SEP-2000; 2000US-0236111P.  
PR 28-SEP-2000; 2000US-0236842P.  
PR 29-SEP-2000; 2000US-0236891P.  
PR 02-OCT-2000; 2000US-0237172P.  
PR 02-OCT-2000; 2000US-0237173P.  
PR 02-OCT-2000; 2000US-0237278P.  
PR 02-OCT-2000; 2000US-0237294P.  
PR 02-OCT-2000; 2000US-0237295P.  
PR 02-OCT-2000; 2000US-0237316P.  
PR 03-OCT-2000; 2000US-0237425P.  
PR 03-OCT-2000; 2000US-0237598P.  
PR 03-OCT-2000; 2000US-0237604P.  
PR 03-OCT-2000; 2000US-0237606P.  
PR 03-OCT-2000; 2000US-0237608P.  
PR 01-NOV-2000; 2000US-0244867P.  
PR 01-NOV-2000; 2000US-0245084P.  
PA (AVAL-) AVALON PHARM.  
PI Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;  
PI Soppet DR, Weaver Z;  
XX WPI; 2002-188264/24.  
XX Screening for anti-neoplastic agent involves exposing cells to a chemical  
XX agent to be tested for anti-neoplastic activity, and determining a change  
XX in expression of a gene of a signature gene set.  
XX Claim 1; SEQ ID NO 4272; 44pp; English.  
XX The present invention describes a method (M1) for screening for an anti-  
XX neoplastic agent. The method involves exposing cells to a chemical agent  
XX to be tested for anti-neoplastic activity, determining a change in  
XX expression of at least one gene (I) of a signature gene set, where (I)  
XX comprises a sequence (S) selected from 847 sequences (given in AB61664  
XX to ABU70110), or is at least 95% identical to (S), where a change in  
XX expression is indicative of anti-neoplastic activity. (I) has cytostatic  
XX activity and can be used in gene therapy. M1 can be used for screening an  
XX anti-neoplastic agent, and can be used for producing a product which is  
XX the data collected with respect to the anti-neoplastic agent as a result  
XX of M1, and the data is sufficient to convey the chemical structure and/or  
XX properties of the agent. M1 can be used in the treatment of cancer such  
XX as colon, breast, stomach, lung, thyroid, oesophageal, ovarian, kidney,  
XX prostate or pancreatic cancer, adenocarcinoma, carcinoma, clear cell  
XX cancer, infiltrating ductal cancer, infiltrating lobular cancer, squamous  
XX cell carcinoma, neuroendocrine carcinoma, papillary carcinoma and Wilm's  
XX tumour  
XX Sequence 697 BP; 190 A; 164 C; 156 G; 180 T; 0 U; 7 Other;  
XX Query Match 45.7%; Score 581; DB 6; Length 697;  
XX Best Local Similarity 95.1%; Pred. No. 2.8e-124;  
XX Matches 621; Conservative 0; Mismatches 27; Indels 5; Gaps 2;  
XX 620 TGACACGCTCAGCTGGGCTGAGTACATCGCCAGTGGAGGAGATCCTGGGTACG 679  
XX 1 TGACACGCTCAGCTGGGCTGAGTACATCGCCAGTGGAGGAGATCCTGGGTACG 60  
XX 680 ACAATACGAGACGGGTATCACTCCCGCTCAACCTGACGTGGCCCTTTATGGTGGCG 739

Db 61 ACAATACGAGAACGGGTACATTCACCCGGTCAACCTGAGTGGCCCTTTATGGTGGCG 120  
Qy 740 GGAACCCGAGAGTGACCTGAAAGAGTGGTACCCGAGCGCCTTATGTGAACACCG 799  
Db 121 GGAACCCGAGAGTGACCTGAAAGAGTGGTACCCGAGCGCCTTATGTGAACACCG 180  
Qy 800 GTCCTGACCTGGAGTGGAGTGGGAAAGCGCGCTCCGGGGGA-GGGGGCCC 858  
Db 181 CGTCCTGACCTGGAGTGGAGTGGGAAAGCGCGCTCCGGGGGANGCCNCNT 240  
Qy 859 GGAAGGCGACCCCTGCGCTCAGTGTCTCTGTCTCTGTCTCTCCCTCGCAATGCTCCTC 918  
Db 241 GGAAGGCGACCCCTGCGCTCAGTGTCTCTGTCTCTGTCTCTCCCTCGCAATGCTCCTC 300  
Qy 919 TCTCTGTCCACCCGCGGAGAACACTTACACGAGGAGAGATTCGTTCCAAACAGA 978  
Db 301 TCTCTGTCCACCCGCGGAGAACACTTACACGAGGAGAGATTCGTTCCAAACAGA 360  
Qy 979 GGAGATCAATTGTACTTACAAAGATCCCATCTATTAACTTTATTAACTTCTACCGTGA 1038  
Db 361 GGAGATCAATTGTACTTACAAAGATTCGTTCCAAACAGA 420  
Qy 1039 ATGACTCTGCAAGCCTTGGTGGTCCAAGTGAATATGTAATATAATATAATAGAT 1098  
Db 421 ATGACTCTGCAAGCCTTGGTGGTCCAAGTGAATATGTAATATAATATAATAGAT 480  
Qy 1099 AAGAGCCTATCAATGATATCTTTGTACATATCTTGTAAATGTAGATCATAGTAGCT 1158  
Db 481 AAGAGCCTATCAATGATATCTTTGTACATATCTTGTAAATGTAGATCATAGTAGCT 540  
Qy 1159 GACTTTGACAGTCAATTTATAAAGTAATCACTTAAAGATATATATTTTTTCAACAA 1218  
Db 541 GACTTTGACAGTCAATTTATAAAGTAATCACTTAAAGATATATATTTTTTCAACAA 597  
Qy 1219 GTTTTGTACTTTTGAATAAATCTTTCTTTATATGCTTAAATAAAAAA 1271  
Db 598 -AGTTTGCACCTTTGAAATAAACCTTTTATATGCTTAAATAAAAAAAGA 649  
XX RESULT 4  
XX ABK64784  
XX ID ABK64784 standard; DNA; 697 BP.  
XX AC ABK64784;  
XX XX  
XX 18-JUN-2002 (first entry)  
XX DE Human benign prostatic hyperplasia gene #679.  
XX XX  
XX Human; benign prostatic hyperplasia; BPH; prostate cancer; gene; ds.  
XX OS Homo sapiens.  
XX PN WO200212440-A2.  
XX PD 14-FEB-2002.  
XX PF 07-AUG-2001; 2001WO-US024708.  
XX PR 07-AUG-2000; 2000US-0223323P.  
XX PR 05-JUN-2001; 2001US-00873319.  
XX PA (GENE-) GENE LOGIC INC.  
XX PI (NIS) JAPAN TOBACCO INC.  
XX PI Munger WE, Kulkarni P, Getzenberg RH, Waga I, Yamamoto J;  
XX WPI; 2002-257476/30.  
XX Identifying drugs for and diagnosing benign prostatic hyperplasia, by  
XX detecting expression levels of one or more genes in prostate cells from  
XX patient that are differentially regulated compared to normal prostate

PT cells.  
 XX Disclosure, Page 367; 444pp; English.  
 PS  
 XX The invention relates to a method of diagnosing (I) the onset or progression of benign prostatic hyperplasia (BPH), or screening (II) for or identifying an agent that modulates the onset or progression of BPH. The method is based on changes in gene expression in BPH tissue isolated from patients exhibiting different clinical states of prostate hyperplasia as compared to normal prostate tissue. (I) comprises detecting the expression levels of one or more genes in prostate cells from the subject that are differentially regulated compared to normal prostate cells. (II) comprises preparing a first gene expression profile of BPH cells or BPH-like cell population, exposing the cells to the agent, preparing a second gene expression profile of the agent exposed cells, and comparing the first and second gene expression profiles. (I) is useful for diagnosing the onset or progression of BPH. (II) is useful for identifying an agent that modulates the onset or progression of BPH. The methods are useful to present information identifying the expression level in a tissue or cells, by comparing the expression level of genes given in the specification in the tissue or cells to the level of expression of gene in the database, and displaying the expression levels of at least one gene in the tissue or cell sample compared to the expression level in BPH. Agents using (II) are useful for treating BPH or prostate cancer. ABK64106-ABK64860 represent human benign prostatic hyperplasia gene sequences of the invention  
 XX  
 SQ Sequence 697 BP; 190 A; 164 C; 156 G; 180 T; 0 U; 7 Other;  
 Query Match 45.7%; Score 581; DB 6; Length 697;  
 Best Local Similarity 95.1%; Pred. No. 2.8e-124;  
 Matches 621; Conservative 0; Mismatches 27; Indels 5; Gaps 2;  
 QY 620 TGGACACGCTCAGGCTGGGCTCCAGCTACATGCCACCTTGAGGAGATCCCTGGTAAAG 679  
 Db 1 TGGACACGCTCAGGCTGGGCTCCAGCTACATGCCACCTTGAGGAGATCCCTGGTAAAG 60  
 QY 680 ACAATACGAGACGGGTACATCAACCGGTCAACCTGACGTGGCCCTTTATGGTGGCG 739  
 Db 61 ACAATACGAGACGGGTACATCAACCGGTCAACCTGACGTGGCCCTTTATGGTGGCG 120  
 QY 740 GAAACCCGAGAGTACCTGAAGAAGTGTGACCGGAGCGCTTATGTGAACACCG 799  
 Db 121 GAAACCCGAGAGTACCTGAAGAAGTGTGACCGGAGCGCTTATGTGAACACCG 180  
 QY 800 CFTCTGACCTTGGAGTCCGAGTCTGGGAAAGGCGCGTCCCGGGGGA-GCGGGCCCC 858  
 Db 181 CFTCTGACCTTGGAGTCCGAGTCTGGGAAAGGCGCGTCCCGGGGAGCGGCGCNC 240  
 QY 859 GGAAGGCGACCCCTGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 918  
 Db 241 GGAAGGCGACCCCTGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 300  
 QY 919 TCTCTGTCACCCCGGAGACACTTTACACGACGAGGAGATTCGTTTCCAAACACAGA 978  
 Db 301 TCTCTGTCACCCCGGAGACACTTTACACGACGAGGAGATTCGTTTCCAAACACAGA 360  
 QY 979 GGAGATCAATGTACTTACAAAGATTCACATCTATTAACTTATTAATCTTACCGTGA 1038  
 Db 361 GGAGATCAATGTACTTACAAAGATTCACATCTATTAACTTATTAATCTTACCGTGA 420  
 QY 1039 ATGACTCTCAAGCCTTGGTGGTCAAGTGAATATGTAATATTAATATATATATAGAT 1098  
 Db 421 ATGACTCTCAAGCCTTGGTGGTCAAGTGAATATGTAATATTAATATATATATAGAT 480  
 QY 1099 AAGAGCCTATCAATGTATCTTTGTACAAATATGTTGTAAGATGATAGATAGTAGCT 1158  
 Db 481 AAGAGCCTATCAATGTATCTTTGTACAAATATGTTGTAAGATGATAGATAGTAGCT 540  
 QY 1159 GACTTGGAGTCACATTTATTAAGTAACTTCACTTAAGATATATATTTTTCACAA 1218  
 Db 541 GACTTGGAGTCACATTTATTAAGTAACTTCACTTAAGATATATATTTTTCACAA 597

QY 1219 GTTTTGCTACTTTGAAATAAATCTTTCTTATATGCTTAAATAAATAAATAA 1271  
 Db 598 -AGTTGCACTTTTGAATAAATCTTTCTTATATGCTTAAATAAATAAATAA 649  
 RESULT 5  
 ABQ55427  
 ID ABQ55427 standard; cDNA; 632 BP.  
 XX  
 AC ABQ55427;  
 XX  
 DT 22-AUG-2002 (first entry)  
 XX  
 DE Human ovarian antigen HNOJ185 cDNA, SEQ ID NO:1307.  
 XX  
 KW Human; ovarian antigen; ovary; ovarian; breast; cancer; tumour;  
 KW ovarian cancer; breast cancer; tumour; reproductive system disorder;  
 KW infertility; pregnancy disorder; anovulation; polycystic ovary syndrome;  
 KW PCOS; ovarian cyst; dysmenorrhea; endocrine disorder; infection;  
 KW inflammatory condition; immune disorder; blood disorder;  
 KW cardiovascular disorder; respiratory disorder; neurological disorder;  
 KW gastrointestinal disorder; urinary system disorder; drug screening;  
 KW gene therapy; chromosome mapping; forensic analysis;  
 KW antibody preparation; cytostatic; immunomodulatory; neuroprotective;  
 KW antiinflammatory; gynaecological; reproductive; gene; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200200677-A1.  
 XX  
 PD 03-JAN-2002.  
 XX  
 PF 07-JUN-2001; 2001WO-US018569.  
 XX  
 PR 07-JUN-2000; 2000US-0209467P.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 PI Birse CE, Rosen CA;  
 XX WPI; 2002-147878/19.  
 DR P-FSDB; ABP42350.  
 XX  
 Isolated nucleic acid molecules encoding novel ovarian polypeptides, useful in the prevention, treatment and diagnosis of cancer (e.g. ovarian cancer), immune disorders, cardiovascular disorders and neurological diseases.  
 PS  
 Claim 1; SEQ ID NO 1307; 2922pp; English.  
 XX  
 The invention relates to 2175 novel human ovarian antigens (ABP41054-ABP43228) and to cDNAs encoding them (ABQ54131-ABQ56305), and also encompasses polypeptides 90% identical and polynucleotides 95% identical to the sequences of the invention. The invention additionally relates to recombinant vectors and host cells comprising human ovarian antigen polynucleotides, antibodies against human ovarian antigens, and the use of ovarian antigen polynucleotides and polypeptides in diagnosing, treating, prognosing or preventing various ovary and/or breast-related disorders. Such conditions include ovarian cancer and breast cancer, and metastatic tumours of ovarian or breast origin, reproductive system disorders (e.g., infertility, disorders of pregnancy, anovulation, polycystic ovary syndrome, ovarian cysts, and dysmenorrhoea), endocrine disorders, infections (e.g., chlamydia, HIV, toxoplasmosis, and toxic shock syndrome), inflammatory conditions (e.g., mastitis, oophoritis and vaginitis), immune disorders (e.g., congenital and acquired immunodeficiencies, autoimmune oophoritis, systemic lupus erythematosus), blood-related disorders (e.g., anaemia), cardiovascular disorders, respiratory disorders, neurological disorders, gastrointestinal disorders and urinary system disorders. Ovarian antigen polypeptides and polynucleotides may also be used in screening for compounds which modulate ovarian antigen expression or activity. The polynucleotides may further be used for gene therapy, chromosome mapping, in the identification of individuals and in forensic analysis, and the

CC polypeptides may be used as food additives or to prepare antibodies  
 CC useful in disease diagnosis, drug targeting and phenotyping. The present  
 CC sequence represents cDNA encoding a human ovarian antigen of the  
 CC invention. Note: The sequence data for this patent did not form part of  
 CC the printed specification, but was obtained in electronic format directly  
 CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

XX SQ Sequence 632 BP; 145 A; 176 C; 201 G; 93 T; 0 U; 17 Other;

Query Match 41.3%; Score 525.6; DB 6; Length 632;  
 Best Local Similarity 93.4%; Pred. No. 1.8e-111;  
 Matches 593; Conservative 0; Mismatches 36; Indels 6; Gaps 5;

QY 257 CTCCTCCCAACATGTCCACGGCTCCCTCAGCGATGTGGAGGACCTTCAAGAGGTGGAGA 316  
 DB 2 CTCCTCCCAACATGTCCACGGCTCCCTCAGCGATGTGGAGGACCTTCAAGAGGTGGAGA 61

QY 317 TGTGGAATGTGACGGGTTGAAATGGAATTCGAACAAGGAATTTGTGACTTCCAAACGAGA 376  
 DB 62 TGTGGAATGTGACGGGTTGAAATGGAATTCGAACAAGGAATTTGTGACTTCCAAACGAGA 121

QY 377 GCACCGAGAGAGCTCAACTGGGAGAAATGGGTCTCCCGAGAAAGGCGCGCGGCGCTGG 436  
 DB 122 GCACCGAGAGAGCTCAACTGGGAGAAATGGGTCTCCCGAGAAAGGCGCGCGGCGCTGG 181

QY 437 GCAAGAGGAGGAGGCGGCCCAACCAAGAGAGCGCCCTGAGCGGGTTCAGCCAGGAGGGGA 496  
 DB 182 GCAAGAGGAGGAGGCGGCCCAACCAAGAGAGCGCCCTGAGCGGGTTCAGCCAGGAGGGGA 241

QY 497 AGCAGTCCAGCGCAACCGCGCCCAACCGCGAGAGCGGCGCCGATGGAGTGTGAGCA 556  
 DB 242 AGCAGTCCAGCGCAACCGCGCCCAACCGCGAGAGCGGCGCCGATGGAGTGTGAGCA 301

QY 557 AGSCCTTCTCCAGACTCAAGACACACCTCCCTGGTGGTCCCGCCGACCAAGCTCTCCA 616  
 DB 302 AGSCCTTCTCCAGACTCAAGACACACCTCCCTGGTGGTCCCGCCGACCAAGCTCTCCA 361

QY 617 AGCTGACACGCTCAGGCTGGGCTCAGTACATCGCCCACTTGAGGAGATCTCTGGTA 676  
 DB 362 AGCTGACACGCTCAGGCTGGGCTCAGTACATCGCCCACTTGAGGAGATCTCTGGTA 421

QY 677 ACACAAATACGAGAGCGGTACATTCACCGGTCACTGACGTGGCCCTTATGTGGTGG 736  
 DB 422 ACACAAATACGAGAGCGGTACATTCACCGGTCACTGACGTGGCCCTTATGTGGTGG 479

QY 737 CCGGGAACCCGAGAGTGACCTGAAAGAGTGG-TGACCGGAGCGCGCTTATGTGGAGCC 795  
 DB 480 -CGGGAACCCGAGAGTGAGTGAAGAGTGGTTGACCGGAGCGCGTATGTGGAANC 538

QY 796 ACCGCTCTGACCTTGGAGGTGCGAGTCTGGAAAGGCGCGCTCCCGGGGGAGCGGCG 855  
 DB 539 ACCGCTCTGACCTTGGAGGTGCGAGTCTGGAAAGGCGCGCTCCCGGGGGAGCGGCG 596

QY 856 CCGGGAAGCGGAGCGGCTCCCTCAGTCTCTCTG 890  
 DB 597 CCGGGAAGCGGAGCGGCTCCCTCAGTCTCTCTG 631

RESULT 6  
 ABL66121/c  
 ID ABL66121 standard; DNA; 471 BP.  
 AC ABL66121;  
 XX  
 DT 15-MAY-2002 (first entry)  
 DE Lung cancer related gene sequence SEQ ID NO:4458.  
 XX Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;  
 KW stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;  
 KW cytostatic; gene therapy; antineoplastic; Wilms tumour; adenocarcinoma;  
 KW gene; ds.  
 XX

OS Homo sapiens.  
 XX WO200194629-A2.  
 XX 13-DEC-2001.  
 XX PF 30-MAY-2001; 2001WO-US010838.  
 XX 05-JUN-2000; 2000US-0209473P.  
 PR 05-JUN-2000; 2000US-0209531P.  
 PR 18-SEP-2000; 2000US-0233133P.  
 PR 18-SEP-2000; 2000US-0233167P.  
 PR 20-SEP-2000; 2000US-0234009P.  
 PR 20-SEP-2000; 2000US-0234034P.  
 PR 20-SEP-2000; 2000US-0234052P.  
 PR 22-SEP-2000; 2000US-0234509P.  
 PR 22-SEP-2000; 2000US-0234567P.  
 PR 25-SEP-2000; 2000US-0234923P.  
 PR 25-SEP-2000; 2000US-0234924P.  
 PR 25-SEP-2000; 2000US-0235077P.  
 PR 25-SEP-2000; 2000US-0235082P.  
 PR 25-SEP-2000; 2000US-0235134P.  
 PR 25-SEP-2000; 2000US-0235280P.  
 PR 26-SEP-2000; 2000US-0235637P.  
 PR 26-SEP-2000; 2000US-0235638P.  
 PR 27-SEP-2000; 2000US-0235711P.  
 PR 27-SEP-2000; 2000US-0235720P.  
 PR 27-SEP-2000; 2000US-0235840P.  
 PR 27-SEP-2000; 2000US-0235863P.  
 PR 28-SEP-2000; 2000US-0236028P.  
 PR 28-SEP-2000; 2000US-0236032P.  
 PR 28-SEP-2000; 2000US-0236033P.  
 PR 28-SEP-2000; 2000US-0236034P.  
 PR 28-SEP-2000; 2000US-0236109P.  
 PR 28-SEP-2000; 2000US-0236111P.  
 PR 29-SEP-2000; 2000US-0236842P.  
 PR 29-SEP-2000; 2000US-0236891P.  
 PR 02-OCT-2000; 2000US-0237172P.  
 PR 02-OCT-2000; 2000US-0237173P.  
 PR 02-OCT-2000; 2000US-0237278P.  
 PR 02-OCT-2000; 2000US-0237294P.  
 PR 02-OCT-2000; 2000US-0237295P.  
 PR 02-OCT-2000; 2000US-0237316P.  
 PR 03-OCT-2000; 2000US-0237425P.  
 PR 03-OCT-2000; 2000US-0237598P.  
 PR 03-OCT-2000; 2000US-0237604P.  
 PR 03-OCT-2000; 2000US-0237606P.  
 PR 03-OCT-2000; 2000US-0237608P.  
 PR 01-NOV-2000; 2000US-0244867P.  
 PR 01-NOV-2000; 2000US-0245084P.  
 XX  
 PA (AVAL-) AVALON PHARM.  
 XX Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;  
 PI Soppet DR, Weaver Z;  
 XX WPI; 2002-188264/24.  
 XX

Screening for anti-neoplastic agent involves exposing cells to a chemical agent to be tested for anti-neoplastic activity, and determining a change in expression of a gene of a signature gene set.

Claim 1; SEQ ID NO 4458; 44pp; English.

The present invention describes a method (M1) for screening for an anti-neoplastic agent. The method involves exposing cells to a chemical agent to be tested for anti-neoplastic activity, determining a change in expression of at least one gene (I) of a signature gene set, where (I) comprises a sequence (S) selected from 8447 sequences (given in ABL61664 to ABL70110), or is at least 95% identical to (S), where a change in expression is indicative of anti-neoplastic activity. (I) has cytostatic activity and can be used in gene therapy. M1 can be used for screening an anti-neoplastic agent, and can be used for producing a product which is





Query Match 36.7%; Score 466.2; DB 6; Length 471;  
 Best Local Similarity 99.4%; Pred. No. 8.7e-98;  
 Matches 468; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 789 TGAACACCGCGTCTGACCTTGAGGTCGAGTCTGGGAAGCGCGCTCCCGGGGG 848  
 DB 471 TGAACACCGCGTCTGACCTTGAGGTCGAGTCTGGGAAGCGCGCTCCCGGGGG 412

QY 849 AGCGGCCCGGGAAGGAGCGACCCCTGCCCTCAGTCTCTGCTCTGCTCCCGCTCGC 908  
 DB 411 AGCGGCCCGGGAAGGAGCGACCCCTGCCCTCAGTCTCTGCTCTGCTCCCGCTCGC 352

QY 909 AATGCTCCCTCTCTGCTCCCGGAGGAGCGACCTTTACACGACGAGGAGATTCGTTT 968  
 DB 351 AATGCTCCCTCTCTGCTCCCGGAGGAGCGACCTTTACACGACGAGGAGATTCGTTT 292

QY 969 CCAAAACGAGGAGATCAATTGTACTTACAAAGATTCCTTAACTTTTAACT 1028  
 DB 291 CCAAAACGAGGAGATCAATTGTACTTACAAAGATTCCTTAACTTTTAACT 232

QY 1029 TCTACCGTGAATGACTCTGCAAGCGCTTGTGTCCTCAAGTGCATATGTAATTAATAT 1088  
 DB 231 TCTACCGTGAATGACTCTGCAAGCGCTTGTGTCCTCAAGTGCATATGTAATTAATAT 172

QY 1089 ATAAATAGATAAGAGCTATCAATGATCTTTTGTACAATATGTTGTAATATGATCA 1148  
 DB 171 ATAAATAGATAAGAGCTATCAATGATCTTTTGTACAATATGTTGTAATATGATCA 112

QY 1149 TAGATAGCTGCTTTGACAGTCACATTTTAAAGTAATTCATTTAAAGATATATTTT 1208  
 DB 111 TAGATAGCTGCTTTGACAGTCACATTTTAAAGTAATTCATTTAAAGATATATTTT 52

QY 1209 TTTCAACAAGTTTGTCTACTTTTGAATAAATCTTTCTTTATATGCTA 1259  
 DB 51 TTTCAACAAGTTTGTCTACTTTTGAATAAATCTTTCTTTATATGCTA 1

RESULT 8  
 ABK64115/c  
 ID ABK64115 standard; DNA; 471 BP.  
 AC ABK64115;  
 XX  
 DT 18-JUN-2002 (first entry)  
 XX  
 DE Human benign prostatic hyperplasia gene #10.  
 XX  
 KW Human; benign prostatic hyperplasia; BPH; prostate cancer; gene; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 PN W0200212440-A2.  
 XX  
 PD 14-FEB-2002.  
 XX  
 PF 07-AUG-2001; 2001WO-US024708.  
 XX  
 PR 07-AUG-2000; 2000US-022323P.  
 PR 05-JUN-2001; 2001US-00873319.  
 XX  
 PA (GENE-) GENE LOGIC INC.  
 PA (NISE ) JAPAN TOBACCO INC.  
 XX  
 PI Munger WE, Kulkarni P, Getzenberg RH, Waga I, Yamamoto J;  
 XX  
 DR WPI; 2002-257476/30.  
 XX  
 PT Identifying drugs for and diagnosing benign prostatic hyperplasia, by  
 PT detecting expression levels of one or more genes in prostate cells from  
 PT patient that are differentially regulated compared to normal prostate  
 PT cells.  
 XX

PS Disclosure; Page 77; 444pp; English.  
 XX The invention relates to a method of diagnosing (I) the onset or  
 CC progression of benign prostatic hyperplasia (BPH), or screening (II) for  
 CC or identifying an agent that modulates the onset or progression of BPH.  
 CC The method is based on changes in gene expression in BPH tissue isolated  
 CC from patients exhibiting different clinical states of prostate  
 CC hyperplasia as compared to normal prostate tissue. (I) comprises  
 CC detecting the expression levels of one or more genes in prostate cells  
 CC from the subject that are differentially regulated compared to normal  
 CC prostate cells. (II) comprises preparing a first gene expression profile  
 CC of BPH cells or BPH-like cell population, exposing the cells to the  
 CC agent, preparing a second gene expression profile of the agent exposed  
 CC cells, and comparing the first and second gene expression profiles. (I)  
 CC is useful for diagnosing the onset or progression of BPH. (II) is useful  
 CC for identifying an agent that modulates the onset or progression of BPH.  
 CC The methods are useful to present information identifying the expression  
 CC level in a tissue or cells, by comparing the expression level of genes  
 CC given in the specification in the tissue or cells to the level of  
 CC expression of gene in the database, and displaying the expression levels  
 CC of at least one gene in the tissue or cell sample compared to the  
 CC expression level in BPH. Agents using (II) are useful for treating BPH or  
 CC prostate cancer. ABK64106-ABK64860 represent human benign prostatic  
 CC hyperplasia gene sequences of the invention  
 XX

QY Sequence 471 BP; 142 A; 89 C; 110 G; 130 T; 0 U; 0 Other;  
 DB Query Match 36.7%; Score 466.2; DB 6; Length 471;  
 Best Local Similarity 99.4%; Pred. No. 8.7e-98;  
 Matches 468; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 789 TGAACACCGCGTCTGACCTTGAGGTCGAGTCTGGGAAGCGCGCTCCCGGGGG 848  
 DB 471 TGAACACCGCGTCTGACCTTGAGGTCGAGTCTGGGAAGCGCGCTCCCGGGGG 412

QY 849 AGCGGCCCGGGAAGGAGCGACCCCTGCCCTCAGTCTCTGCTCTGCTCCCGCTCGC 908  
 DB 411 AGCGGCCCGGGAAGGAGCGACCCCTGCCCTCAGTCTCTGCTCTGCTCCCGCTCGC 352

QY 909 AATGCTCCCTCTCTGCTCCCGGAGGAGCGACCTTTACACGACGAGGAGATTCGTTT 968  
 DB 351 AATGCTCCCTCTCTGCTCCCGGAGGAGCGACCTTTACACGACGAGGAGATTCGTTT 292

QY 969 CCAAAACGAGGAGATCAATTGTACTTACAAAGATTCCTTAACTTTTAACT 1028  
 DB 291 CCAAAACGAGGAGATCAATTGTACTTACAAAGATTCCTTAACTTTTAACT 232

QY 1029 TCTACCGTGAATGACTCTGCAAGCGCTTGTGTCCTCAAGTGCATATGTAATTAATAT 1088  
 DB 231 TCTACCGTGAATGACTCTGCAAGCGCTTGTGTCCTCAAGTGCATATGTAATTAATAT 172

QY 1089 ATAAATAGATAAGAGCTATCAATGATCTTTTGTACAATATGTTGTAATATGATCA 1148  
 DB 171 ATAAATAGATAAGAGCTATCAATGATCTTTTGTACAATATGTTGTAATATGATCA 112

QY 1149 TAGATAGCTGCTTTGACAGTCACATTTTAAAGTAATTCATTTAAAGATATATTTT 1208  
 DB 111 TAGATAGCTGCTTTGACAGTCACATTTTAAAGTAATTCATTTAAAGATATATTTT 52

QY 1209 TTTCAACAAGTTTGTCTACTTTTGAATAAATCTTTCTTTATATGCTA 1259  
 DB 51 TTTCAACAAGTTTGTCTACTTTTGAATAAATCTTTCTTTATATGCTA 1

RESULT 9  
 ACH23969  
 ID ACH23969 standard; cDNA; 493 BP.  
 XX  
 AC ACH23969;  
 XX  
 DT 13-OCT-2003 (first entry)  
 XX  
 DE Human adult ovary cDNA #2349.

```
XX KW Human; ss; sequencing by hybridisation; SBH; expressed sequence tag; EST;
XX QY genome mapping; biodiversity; genetic disorder.
XX OS Homo sapiens.
XX PN US2003073623-A1.
XX PD 17-APR-2003.
XX PF 30-JUL-2001; 2001US-00918995.
XX PR 30-JUL-2001; 2001US-00918995.
XX PA (DRMA/) DRMANAC R T.
XX PA (LABA/) LABAT I.
XX PA (STAC/) STACHE-CRAIN B.
XX PA (DICK/) DICKSON M C.
XX PA (JONE/) JONES L W.
XX PI Drmanac RT, Labat I, Stache-Crain B, Dickson MC, Jones LW;
XX WPI; 2003-615964/58.
XX New polynucleotide sequences obtained from various cDNA libraries, useful
PT as hybridization probes, as oligomers for PCR, for chromosome and gene
PT mapping, in the recombinant production of protein, or in generating
PT antisense DNA or RNA.
XX Claim 1; SEQ ID NO 11181; 44pp; English.
XX The invention relates to an isolated polynucleotide comprising any one of
CC 38043 cDNA sequences, appearing as ACH12789-ACH50831, whose sequence was
CC determined by the technique of SBH (sequencing by hybridisation). Also
CC included is a purified polypeptide comprising a sequence corresponding to
CC a reading frame of the novel polynucleotide. The nucleic acid sequences
CC are useful in diagnostics as expressed sequence tags (EST) for
CC identifying expressed genes or for physical mapping of the human genome,
CC in forensics, in assessing biodiversity, or in identifying mutations
CC responsible for genetic disorders and other traits. The nucleotide
CC sequences are also useful as hybridisation probes, as oligomers for PCR,
CC for chromosome and gene mapping, in the recombinant production of
CC protein, or in generating antisense DNA or RNA. The purified polypeptide
CC is useful for generating antibodies specific for it. The present sequence
CC is one of the 38043 isolated cDNA/EST sequences. Note: The sequence data
CC for this patent did not form part of the printed specification, but was
CC obtained in electronic format directly from USPTO at
CC seqdata.uspto.gov/sequence.html?DocID=20030073623
XX
XX SQ Sequence 493 BP; 121 A; 136 C; 121 G; 112 T; 0 U; 3 Other;
Query Match 34.1%; Score 434.2; DB 9; Length 493;
Best Local Similarity 99.3%; Pred. No. 2.2e-90;
Matches 436; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 674 CTAACGACAATACGAGACGGGTACATTCACCGGTCAACTGACGTGGCCCTTTATGG 733
DB 55 CTAACGACAATACGAGACGGGTACATTCACCGGTCAACTGATGTGGCCCTTTATGG 114
QY 734 TGGCCGGGAACCCGAGAGTGACCTGAAAGAGTGTGACCGGAGCGCTTATGGAA 793
DB 115 TGGCCGGGAACCCGAGAGTGACCTGAAAGAGTGTGACCGGAGCGCTTATGGAA 174
QY 794 CCACCGCGTCTGACCTTGGAGTGGAGTCTGGGAAAGCGCGTCCCGGGGGAGCGG 853
DB 175 CCACCGCGTCTGACCTTGGAGTGGAGTCTGGGAAAGCGCGTCCCGGGGGAGCGG 234
QY 854 GCCCGGGAAGGACCCCTGCCTCAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 913
DB 235 GCCCGGGAAGGACCCCTGCCTCAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 294
QY 914 TCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 973
```

```
Db 295 TCCTCTCTCTGTCACCCCGGAGACACTTTACACGACGAGGAGATTGGTTCCAAA 354
QY 974 CCAGAGGAGATCAATTCTACTTACAAAGATCCCACTTATTAACTTTATTACTTAC 1033
Db 355 CCAGAGGAGATCAATTCTACTTACAAAGATCCCACTTATTAACTTTATTACTTAC 414
QY 1034 CGTGAATGACTCTGCAAGCCTTGTGTCGCAAGTGCATATGTAATTATAAATATAAA 1093
Db 415 CGTGAATGACTCTGCAAGCCTTGTGTCGCAAGTGCATATGTAATTATAAATATAAA 474
QY 1094 TAGATAAGAGCCTATCAAT 1112
Db 475 TAGATAAGAGCCTATCAAT 493
RESULT 10
ACH21857
ID ACH21857 standard; cDNA; 482 BP.
XX AC ACH21857;
XX DT 13-OCT-2003 (first entry)
XX DE Human adult ovary cDNA #237.
XX KW Human; ss; sequencing by hybridisation; SBH; expressed sequence tag; EST;
XX QY genome mapping; biodiversity; genetic disorder.
XX OS Homo sapiens.
XX PN US2003073623-A1.
XX PD 17-APR-2003.
XX PF 30-JUL-2001; 2001US-00918995.
XX PR 30-JUL-2001; 2001US-00918995.
XX PA (DRMA/) DRMANAC R T.
XX PA (LABA/) LABAT I.
XX PA (STAC/) STACHE-CRAIN B.
XX PA (DICK/) DICKSON M C.
XX PA (JONE/) JONES L W.
XX PI Drmanac RT, Labat I, Stache-Crain B, Dickson MC, Jones LW;
XX WPI; 2003-615964/58.
XX New polynucleotide sequences obtained from various cDNA libraries, useful
PT as hybridization probes, as oligomers for PCR, for chromosome and gene
PT mapping, in the recombinant production of protein, or in generating
PT antisense DNA or RNA.
XX Claim 1; SEQ ID NO 9069; 44pp; English.
XX The invention relates to an isolated polynucleotide comprising any one of
CC 38043 cDNA sequences, appearing as ACH12789-ACH50831, whose sequence was
CC determined by the technique of SBH (sequencing by hybridisation). Also
CC included is a purified polypeptide comprising a sequence corresponding to
CC a reading frame of the novel polynucleotide. The nucleic acid sequences
CC are useful in diagnostics as expressed sequence tags (EST) for
CC identifying expressed genes or for physical mapping of the human genome,
CC in forensics, in assessing biodiversity, or in identifying mutations
CC responsible for genetic disorders and other traits. The nucleotide
CC sequences are also useful as hybridisation probes, as oligomers for PCR,
CC for chromosome and gene mapping, in the recombinant production of
CC protein, or in generating antibodies specific for it. The present sequence
CC is one of the 38043 isolated cDNA/EST sequences. Note: The sequence data
CC for this patent did not form part of the printed specification, but was
CC obtained in electronic format directly from USPTO at
CC seqdata.uspto.gov/sequence.html?DocID=20030073623
XX
```

SQ Sequence 482 BP; 99 A; 156 C; 143 G; 83 T; 0 U; 1 Other;

Query Match 23.4%; Score 297.2; DB 9; Length 482;  
Best Local Similarity 97.1%; Pred. No. 1.1e-58;  
Matches 302; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 517 GCCAAGCGGAGAGCGGGCCGCGATGCTGAGTCTGAGCAAGCCCTTCTCCAGACTCAAG 576  
DB 33 GCNNTAGCGGAGAGCGGGCCGCGATGCTGAGTCTGAGCAAGCCCTTCTCCAGACTCAAG 92

QY 577 ACCACCTGCTGGTGGTGGGCCCCGACACCAAGCTTCCAAAGTGCACAGCTCAGGCTG 636  
DB 93 ACCACCTGCTGGTGGTGGGCCCCGACACCAAGCTTCCAAAGTGCACAGCTCAGGCTG 152

QY 637 GCGTCAGCTACATCGCCACTTGGCGAGATCCTGGCTAACGACAAATACAGAACCGG 696  
DB 153 GCGTCAGCTACATCGCCACTTGGCGAGATCCTGGCTAACGACAAATACAGAACCGG 212

QY 697 TACATTACCCGGTCAACCTGACGTGGCCCTTTATGGTGGCCGGGAAACCCGAGAGTGAC 756  
DB 213 TACATTACCCGGTCAACCTGACGTGGCCCTTTATGGTGGCCGGGAAACCCGAGAGTGAC 272

QY 757 CTGAAGAAGTGTGACCGGAGCCCTTATGTGGAACCAACCGCTGACCTTGGAGG 816  
DB 273 CTGAAGAAGTGTGACCGGAGCCCTTATGTGGAACCAACCGCTGACCTTGGAGT 332

QY 817 TGGGAGTCTGG 827  
DB 333 TTGGTACCTGG 343

RESULT 11  
AAC08260  
ID AAC08260 standard; cDNA; 271 BP.  
AC AAC08260;  
XX  
DT 06-OCT-2000 (first entry)  
DE Human secreted protein 5' EST, SEQ ID NO: 12335.  
XX  
XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;  
KW gene therapy; chromosome mapping; ss.  
XX Homo sapiens.  
XX EP1033401-A2.  
XX  
XX 06-SEP-2000.  
XX  
XX 21-FEB-2000; 2000BP-00200610.  
XX  
XX 26-FEB-1999; 99US-0122487P.  
XX  
XX (GIST ) GENSET.  
XX  
XX Dumas Milne Edwards J, Duclert A, Giordano J;  
XX WPI; 2000-500381/45.  
XX  
XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for  
PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for  
PT diagnostic, forensic, gene therapy and chromosome mapping procedures.  
XX  
XX Claim 1; SEQ ID NO 12335; 71pp + Sequence Listing; English.  
XX  
XX The present sequence is one of a large number of 5' ESTs derived from  
CC mRNAs encoding secreted proteins. No ORF has yet been conclusively  
CC identified within the present sequence. The 5' ESTs were prepared from  
CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST  
CC sequences usually correspond mainly to the 3' untranslated region (UTR)  
CC of the mRNA because they are often obtained from oligo-dT primed cDNA  
CC libraries. Such ESTs are not well suited for isolating cDNA sequences

CC derived from the 5' ends of mRNAs and even in those cases where longer  
CC cDNA sequences have been obtained, the full 5' UTR is rarely included. 5'  
CC ESTs are derived from mRNAs with intact 5' ends and can therefore be used  
CC to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used in  
CC diagnostic, forensic, gene therapy and chromosome mapping procedures.  
CC They are used to obtain upstream regulatory sequences and to design  
CC expression and secretion vectors

XX Sequence 271 BP; 41 A; 97 C; 65 G; 67 T; 0 U; 1 Other;

Query Match 19.4%; Score 246.8; DB 3; Length 271;  
Best Local Similarity 98.4%; Pred. No. 3.9e-47;  
Matches 248; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 TCTACGCCACGACTCTGGAGTGGGGAACAGAGAGCGGTTCTCTGCTGCAGAGTC 60  
DB 19 TCTACGCCACGACTCTGGAGTGGGGAACAGAGAGCGGTTCTCTGCTGCAGAGTC 78

QY 61 CTCGGGTTCTTCTCTCAAACTCTGGAAGGGGAAAGGGTTGTGAGACCCACAGACCC 120  
DB 79 CTCGGGTTCTTCTCTCAAACTCTGGAAGGGGAAAGGGTTGTGAGACCCACAGACCC 138

QY 121 CAATCTCAGTCCAGCAGAGAGTGGCTGCGCCACACTCGGAGGCTCTTGGTTTCAGG 180  
DB 139 CAATCTCAGTCCAGCAGAGAGTGGCTGCGCCACACTCGGAGGCTCTTGGTTTCAGG 198

QY 181 GTCTCTCTGTCTCTCTCACCCCTCTTCTCTGCTGCTTCTCTCTCTCTCTCTCT 240  
DB 199 GTCTCTCTGTCTCTCTCACCCCTCTTCTCTGCTGCTTCTCTCTCTCTCTCTCT 258

QY 241 CTCTCTCCCTCG 252  
DB 259 CCTCTTCTCTCG 270

RESULT 12  
ADK61063/c  
ID ADK61063 standard; DNA; 265 BP.  
XX  
AC ADK61063;  
XX  
DT 06-MAY-2004 (first entry)  
DE Ovarian cancer-related DNA #218 with altered ovarian cancer expression.  
XX  
XX ds; gene; ovarian tumor; BRCA-1-like; BRCA-2-like; non-BRCA-like;  
KW gene expression; primer; cancer.  
XX Homo sapiens.  
XX WO2003068054-A2.  
XX  
XX 21-AUG-2003.  
XX  
XX 13-FEB-2003; 2003WO-US004688.  
XX  
XX 13-FEB-2002; 2002US-0357031P.  
XX  
XX (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
PA (SLOK ) SLOAN KETTERING INST CANCER RES.  
XX  
XX Jazaeri AA, Boyd J, Liu ET;  
XX WPI; 2003-689589/65.  
XX  
XX Classifying an ovarian tumor as a BRCA-1-like or BRCA-2-like or non-BRCA-  
PT like tumor by determining a pattern of expression in the ovarian tumor of  
PT several markers.  
XX  
XX Disclosure; SEQ ID NO 233; 137pp; English.  
XX  
XX The invention relates to a method of classifying an ovarian tumor as a  
CC BRCA-1-like or BRCA-2-like or non-BRCA-like tumor by: (1) determining a

CC pattern of expression in the ovarian tumor of several markers given in  
 CC the specification; and (2) comparing a similarity of the pattern of  
 CC expression of the markers in the ovarian tumor to a pattern of expression  
 CC of the markers in a comparison tissue of a known BRCA-1-like or BRCA-2-  
 CC like or non-BRCA-like tumor. The method is useful for classifying an  
 CC ovarian tumor as a BRCA-1-like or BRCA-2-like or non-BRCA-like tumor.  
 CC This sequence corresponds to an ovarian cancer -related gene having an  
 CC altered pattern of expression in ovarian cancer. (Note: The sequence data  
 CC for this patent did not form part of the printed specification but was  
 CC obtained in electronic format from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences).

XX Sequence 265 BP; 98 A; 33 C; 38 G; 96 T; 0 U; 0 Other;

Query Match 18.6%; Score 236.2; DB 10; Length 265;  
 Best Local Similarity 98.1%; Pred. No. 1.1e-44;  
 Matches 260; Conservative 0; Mismatches 3; Indels 2; Gaps 2;

QY 995 TACAAGATCCCACTATTATTAACCTTCTACCGTG-AATGACTCTGCAAGCC 1053  
 DB 265 TACAAGATCCCACTATTATTAACCTTCTACCGTGAAATGACTCTGCAAGCC 206  
 QY 1054 TTGCTGTCACAGTCCCAATGTAATTAATAATATAGATAAGAGCCTATCAATG 1113  
 DB 205 TTGCTGACCAAGTGCAATGTAATTAATAATATAGATAAGAGCCTATCAATG 146  
 QY 1114 TATCTTTTGTACAATATGTTGTAATAATGATAGATAGCTGACTTTGACAGTCAC 1173  
 DB 145 TATCTTTTGTACAATATGTTGTAATAATGATAGATAGCTGACTTTGACAGTCAC 86  
 QY 1174 ATTATAAGTAATTCACCTTAAGATATATATTTTTCACAAAGT-TTTGCTACTTTT 1232  
 DB 85 ATTATAAGTAATTCACCTTAAGATATATATTTTTCACAAAGT-TTTGCTACTTTT 26  
 QY 1233 GAAATAAATCTTCTTTTATATTGC 1257  
 DB 25 GAAATAAATCTTATTTATATTGC 1

## RESULT 13

ID ADM32307/c  
 XX ADM32307 standard; DNA; 265 BP.  
 XX ADM32307;  
 XX 20-MAY-2004 (first entry)  
 DT Human cancer gene, SEQ ID NO 2.  
 XX detection; cancer; cytotoxic; anticancer; cytostatic; gene therapy; ds;  
 XX human.  
 XX Homo sapiens.  
 XX JP2004049122-A.  
 XX 19-FEB-2004.  
 XX 19-JUL-2002; 2002JP-00211830.  
 XX 19-JUL-2002; 2002JP-00211830.  
 XX (CHUS) CHUGAI PHARM CO LTD.  
 XX (GANK-) ZH GAN KENYUKAI.  
 XX WPI; 2004-232131/22.  
 XX Detecting cancer, comprises measuring the expression level of cancer  
 XX related polypeptides selected from a set of polypeptides given in the  
 XX specification.  
 XX Claim 1; SEQ ID NO 2; 116pp; Japanese.

CC The invention relates to a novel method for detecting cancer. The novel  
 CC method involves measuring the expression level of a polypeptide encoded  
 CC by a DNA having a sequence chosen from 119 fully defined sequences such  
 CC as 416, 265, 450, 376, 360, 245, 290, 268, 221, 391 nucleotides etc., as  
 CC given in the specification, or a DNA that hybridizes to the above DNA  
 CC under stringent conditions. The invention further relates to: a  
 CC diagnostic test of cancer; a DNA which encodes a cytotoxic protein and is  
 CC coupled with a functional promoter region of a cancer detecting DNA; a  
 CC vector having cytotoxic DNA inserted in it; a transformed cell containing  
 CC the said vector; a method for evaluating the presence or absence of  
 CC anticancer activity in a test sample; and producing a composition, by  
 CC mixing a sample evaluated by the anticancer activity detection method and  
 CC a carrier. The novel polynucleotides have cytostatic activity. The vector  
 CC containing a cytotoxic polynucleotide can be used to treat cancer by gene  
 CC therapy. The novel method of the invention allows early detection of  
 CC cancer. This polynucleotide represents one of the 119 DNA sequences  
 CC useful in the method for detecting cancer of the invention.

XX Sequence 265 BP; 98 A; 33 C; 38 G; 96 T; 0 U; 0 Other;

Query Match 18.6%; Score 236.2; DB 12; Length 265;  
 Best Local Similarity 98.1%; Pred. No. 1.1e-44;  
 Matches 260; Conservative 0; Mismatches 3; Indels 2; Gaps 2;

QY 995 TACAAGATCCCACTATTATTAACCTTCTACCGTG-AATGACTCTGCAAGCC 1053  
 DB 265 TACAAGATCCCACTATTATTAACCTTCTACCGTGAAATGACTCTGCAAGCC 206  
 QY 1054 TTGCTGTCACAGTCCCAATGTAATTAATAATATAGATAAGAGCCTATCAATG 1113  
 DB 205 TTGCTGACCAAGTGCAATGTAATTAATAATATAGATAAGAGCCTATCAATG 146  
 QY 1114 TATCTTTTGTACAATATGTTGTAATAATGATAGATAGCTGACTTTGACAGTCAC 1173  
 DB 145 TATCTTTTGTACAATATGTTGTAATAATGATAGATAGCTGACTTTGACAGTCAC 86  
 QY 1174 ATTATAAGTAATTCACCTTAAGATATATATTTTTCACAAAGT-TTTGCTACTTTT 1232  
 DB 85 ATTATAAGTAATTCACCTTAAGATATATATTTTTCACAAAGT-TTTGCTACTTTT 26  
 QY 1233 GAAATAAATCTTCTTTTATATTGC 1257  
 DB 25 GAAATAAATCTTATTTATATTGC 1

## RESULT 14

ID ACA56430  
 XX ACA56430 standard; cDNA; 240 BP.  
 XX ACA56430;  
 XX 06-JUN-2003 (first entry)  
 DT Chicken signalling pathway polynucleotide probe SEQ ID NO 1028.  
 XX  
 XX Chicken; probe; ss; array element; Parkinson's disease;  
 XX signalling pathway population; cancer; adenocarcinoma; leukaemia;  
 XX immunopathy; AIDS; asthma; neuropathy; Alzheimer's disease; microarray.  
 XX Gallus gallus.  
 XX US6500938-B1.  
 XX 31-DEC-2002.  
 XX 30-JAN-1998; 98US-00016434.  
 XX 30-JAN-1998; 98US-00016434.  
 XX (INCY-) INCYTE GENOMICS INC.  
 XX Au-Young J, Seilhamer JJ;

1026c and cdp

DR WPI; 2003-352189/33.  
 XX Combination of polynucleotide probes, useful as array elements in a  
 PT microarray for monitoring the expression of a number of target  
 PT polynucleotides.  
 XX  
 PS Claim 1; SEQ ID NO 1028; 65pp; English.  
 XX  
 XX The invention relates to a combination which, comprises a number of  
 CC polynucleotide probes comprising a sequence selected from one of the 1490  
 CC sequences mentioned in the specification. The combination is useful as an  
 CC array element in a microarray for monitoring the expression of a number  
 CC of target polynucleotides. The microarray is particularly useful in the  
 CC diagnosis and treatment of cancer and immunopathology and neuropathology.  
 CC The microarray is useful in diagnostics and treatment regimens, drug  
 CC discovery and development, toxicological and carcinogenicity studies,  
 CC forensics and pharmacogenomics. The microarray is also useful for  
 CC monitoring progression of diseases and for developing sophisticated  
 CC profiles for the effects of currently available therapeutic drugs. The  
 CC combination is also useful for purifying a subpopulation of mRNAs, cDNAs  
 CC and genomic fragments and in research and diagnostic applications. The  
 CC array can detect changes in expression in a large number of genes coding  
 CC for different signaling pathway populations which can be used to diagnose  
 CC various diseases including cancer e.g. adenocarcinoma and leukaemia,  
 CC immunopathies e.g. AIDS and asthma. neuropathies e.g. Alzheimer's disease  
 CC and Parkinson's disease. The present sequence represents a polynucleotide  
 CC probe of the invention. Note: The sequence data for this patent did not  
 CC form part of the printed specification but was obtained in electronic  
 CC format directly from USPTO at  
 CC seqdata.uspto.gov/sequence.html?DocID=0650093831  
 XX  
 SQ Sequence 240 BP; 55 A; 83 C; 72 G; 27 T; 0 U; 3 Other;  
 Query Match 16.7%; Score 213; DB 10; Length 240;  
 Best Local Similarity 97.9%; Pred. No. 2.5e-39;  
 Matches 235; Conservative 0; Mismatches 3; Indels 2; Gaps 2;  
 QY 448 AAGGCGCCCAACGAAGAGCCCTGTAGCGGGGTGAGGCA-GGAGGGGAAGCAGGTCCA 506  
 Db 1 AAGGCGCCCAACGAAGAGCCCTGTAGCGGGGTGAGGCA-GGAGGGGAAGCAGGTCCA 60  
 QY 507 GCGCAACGCGCCCAACGCGCGAGAGC-GGGCCCGCATGCTGAGCAAGGCGCTTCT 565  
 Db 61 GCGCAACGCGCCCAACGCGCGAGAGCGGGCGCGCATGCTGAGCAAGGCGCTTCT 120  
 QY 566 CCAGACTCAAGACCAACCTGCGCTGCGTCCCGCCGACACCAAGCTCTCCAGCTGGACA 625  
 Db 121 CCAGACTCAAGACCAACCTGCGCTGCGTCCCGCCGACACCAAGCTCTCCAGCTGGACA 180  
 QY 626 CGCTCAGGCTGGCGTCCAGTACATCGCCCACTTGAGGCAGATCTCCAGCTGGACA 685  
 Db 181 CGCTCAGGCTGGCGTCCAGTACATCGCCCACTTGAGGCAGATCTCCAGCTGGACA 240  
 RESULT 15  
 ADI56226  
 ID ADI56226 standard; DNA; 240 BP.  
 XX  
 AC ADI56226;  
 XX  
 XX ADI56226;  
 XX  
 DT 22-APR-2004 (first entry)  
 XX  
 DE Human polynucleotide probe #1028.  
 XX  
 KW Human; probe; ss; receptor-like polypeptide; transducing polypeptide;  
 KW effector-like polypeptide; cancer; immunopathology; neuropathology;  
 KW drug development; toxicology; carcinogenicity; immunopathology;  
 KW signalling pathway polypeptide; adrenal gland; bladder; bone;  
 KW bone marrow; brain; breast; cervix; tumour; immunopathology; AIDS;  
 KW diabetes; pancreatitis; osteoporosis; ulcerative colitis; neuropathology;  
 KW dementia; amnesia; epilepsy; Alzheimer's disease; depression.  
 XX  
 OS Homo sapiens.

XX PN US2004010136-A1.  
 XX PD 15-JAN-2004.  
 XX PF 26-NOV-2002; 2002US-00305720.  
 XX PR 30-JAN-1998; 98US-00016434.  
 XX (INCY-) INCYTE GENOMICS INC.  
 XX Au-Young J, Seilhamer JJ;  
 XX WPI; 2004-090520/09.  
 XX  
 PT New composition comprising polynucleotide probes, useful as array  
 PT elements in a microarray for monitoring the expression of target  
 PT polynucleotides or purifying a subpopulation of mRNAs, cDNA, or genomic  
 PT fragments.  
 XX  
 PS Claim 6; SEQ ID NO 1028; 73pp; English.  
 XX  
 XX The invention relates to a composition of polynucleotide probes  
 CC comprising first polynucleotide probes comprising at least a portion of a  
 CC gene encoding a receptor-like polypeptide, second polynucleotide probes  
 CC comprising at least a portion of a gene encoding a transducing  
 CC polypeptide and third polynucleotide probes comprising at least a portion  
 CC of a gene encoding an effector-like polypeptide. The probes of the  
 CC composition are useful as array elements in a microarray for monitoring  
 CC the expression of target polynucleotides. The microarray is useful in the  
 CC diagnosis and treatment of cancer, an immunopathology or a  
 CC neuropathology. It can also be used for drug discovery and development.  
 CC toxicological and carcinogenicity studies, forensics or pharmacogenomics.  
 CC Microarrays can also be used for monitoring the progression of diseases  
 CC that may be associated with the altered expression of signalling pathway  
 CC polypeptides. The composition can also be used to purify a subpopulation  
 CC of mRNAs, cDNAs, or genomic fragments in a sample. The expression profile  
 CC is also useful for the diagnosis and treatment of cancer, e.g. cancers of  
 CC the adrenal gland, bladder, bone, bone marrow, brain, breast or cervix,  
 CC an immunopathology, e.g. AIDS, diabetes, pancreatitis, osteoporosis or  
 CC ulcerative colitis, or a neuropathology, e.g. dementia, amnesia,  
 CC epilepsy, Alzheimer's disease or depression. This sequence represents a  
 CC human polynucleotide probe of the invention. Note: The sequence data for  
 CC this patent did not form part of the printed specification but was  
 CC obtained in electronic format directly from USPTO at  
 CC seqdata.uspto.gov/sequence.html.  
 XX  
 SQ Sequence 240 BP; 55 A; 83 C; 72 G; 27 T; 0 U; 3 Other;  
 Query Match 16.7%; Score 213; DB 12; Length 240;  
 Best Local Similarity 97.9%; Pred. No. 2.5e-39;  
 Matches 235; Conservative 0; Mismatches 3; Indels 2; Gaps 2;  
 QY 448 AAGGCGCCCAACGAAGAGCCCTGTAGCGGGGTGAGGCA-GGAGGGGAAGCAGGTCCA 506  
 Db 1 AAGGCGCCCAACGAAGAGCCCTGTAGCGGGGTGAGGCA-GGAGGGGAAGCAGGTCCA 60  
 QY 507 GCGCAACGCGCCCAACGCGCGAGAGC-GGGCCCGCATGCTGAGCAAGGCGCTTCT 565  
 Db 61 GCGCAACGCGCCCAACGCGCGAGAGCGGGCGCGCATGCTGAGCAAGGCGCTTCT 120  
 QY 566 CCAGACTCAAGACCAACCTGCGCTGCGTCCCGCCGACACCAAGCTCTCCAGCTGGACA 625  
 Db 121 CCAGACTCAAGACCAACCTGCGCTGCGTCCCGCCGACACCAAGCTCTCCAGCTGGACA 180  
 QY 626 CGCTCAGGCTGGCGTCCAGTACATCGCCCACTTGAGGCAGATCTCCAGCTGGACA 685  
 Db 181 CGCTCAGGCTGGCGTCCAGTACATCGCCCACTTGAGGCAGATCTCCAGCTGGACA 240  
 Search completed: December 16, 2004, 17:30:53  
 Job time : 687 secs

Result No.	Score	Query		DB	ID	Description
		Match	Length			
1	246.8	19.4	271	4	US-09-513-999C-12335	Sequence 12335, A
2	213	16.7	240	4	US-09-016-434-1028	Sequence 1028, A
3	71.8	5.6	1800	4	US-09-771-357-106	Sequence 106, A
C 4	64.6	5.1	2337	1	US-07-891-9420-11	Sequence 11, A
C 5	63.4	5.0	305	1	US-08-253-155A-4	Sequence 4, A
6	63.4	5.0	1268	1	US-08-910-973-12	Sequence 12, A
7	63.4	5.0	1268	1	US-09-499-227-12	Sequence 12, A
8	63.2	5.0	3618	3	US-09-042-353-224	Sequence 224, A
9	63.2	5.0	3618	3	US-08-758-417A-72	Sequence 72, A
10	63.2	5.0	3638	1	US-07-634-539A-49	Sequence 49, A
11	63.2	5.0	3638	2	US-08-800-353-49	Sequence 49, A
12	63.2	5.0	3638	5	PCT-US93-06185-49	Sequence 49, A
13	63.2	5.0	3639	1	US-08-953-131-120	Sequence 120, A
14	63.2	5.0	3639	1	US-08-645-641-120	Sequence 120, A
15	63.2	5.0	3699	1	US-07-853-408B-120	Sequence 120, A
16	63.2	5.0	3699	1	US-08-096-762-120	Sequence 120, A
17	63.2	5.0	3699	2	US-08-308-865-120	Sequence 120, A
18	63.2	5.0	3699	5	PCT-US92-10983-120	Sequence 120, A
19	63	5.0	1352	1	US-08-852-142A-10	Sequence 10, A
20	63	5.0	1535	1	US-08-910-973-10	Sequence 10, A
21	63	5.0	1535	4	US-09-499-227-10	Sequence 10, A
22	63	5.0	1550	3	US-09-234-332-3	Sequence 3, A
C 23	62.6	4.9	349	1	US-08-253-155A-14	Sequence 14, A
24	62.2	4.9	165	4	US-09-636-215-736	Sequence 736, A
25	62.2	4.9	165	4	US-09-685-166A-736	Sequence 736, A
26	62.2	4.9	165	4	US-09-679-426-736	Sequence 736, A
27	62	4.9	310	1	US-08-852-142A-12	Sequence 12, A







```
; CLONE: 20A1 (neuroD3)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 55..768
US-08-910-973-12

Query Match
Best Local Similarity 5.0%; Score 63.4; DB 1; Length 1268;
Matches 103; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

QY 504 CCAGCGCAAGCGCGCAAGCGCGAGAGCGGGCGCGCATGCGAGTGTGAGCAAGGCCTT 563
Db 330 CCGCGCGGTCAAGCGCAAGCGATCGCGAGCGCAACCGCATGCACTTGAACCGCGCCT 389
QY 564 CTCGAGCTCAAGACCAACCGTCCCTGGTGGCGCGCGCGCGCGCGCGCGCGCGCG 623
Db 390 GGAGCGACTGCGCGAGCGGTGCGCTCGTTCGCCGAGACCAAGCTCACCAGCTCGA 449
QY 624 CAGCTCAGCTGGCTCGAGTGTACATCGCCCGCGCGCGCGCGCGCGCGCGCGCG 672
Db 450 GAGCGTGGCTTGGCGTGTACATCGCTGCGGTCTGCGCGCGAGACACTG 498

RESULT 7
US-09-499-227-12
; Sequence 12, Application US/09499227
; Patent No. 644463
; GENERAL INFORMATION:
; APPLICANT: Tapscott, Stephen J.
; APPLICANT: Olson, James M.
; TITLE OF INVENTION: Expression of Neurogenic bHLH Genes in Primitive Neuroectoderm
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Christensen O'Connor Johnson Kindness PLLC
; STREET: 1420 Fifth Avenue, Suite 2800
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98101-2347
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/499,227
; FILING DATE: 05-August-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/239,238
; FILING DATE: 06-May-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/US95/05741
; FILING DATE: 08-May-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/17532
; FILING DATE: 30-October-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/910,973
; FILING DATE: 07-August-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Sheiness, Diana K.
; REGISTRATION NUMBER: 35,356
; REFERENCE/DOCKET NUMBER: FHCR-1-12742
; TELEPHONE: 206-682-8100; 206-224-0735 (direct)
; TELEFAX: 206-225-0779
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1268 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
```

```
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; IMMEDIATE SOURCE:
; CLONE: 20A1 (neuroD3)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 55..768
US-09-499-227-12

Query Match
Best Local Similarity 5.0%; Score 63.4; DB 4; Length 1268;
Matches 103; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

QY 504 CCAGCGCAAGCGCGCAAGCGCGAGAGCGGGCGCGCATGCGAGTGTGAGCAAGGCCTT 563
Db 330 CCGCGCGGTCAAGCGCAAGCGATCGCGAGCGCAACCGCATGCACTTGAACCGCGCCT 389
QY 564 CTCGAGCTCAAGACCAACCGTCCCTGGTGGCGCGCGCGCGCGCGCGCGCGCGCG 623
Db 390 GGAGCGACTGCGCGAGCGGTGCGCTCGTTCGCCGAGACCAAGCTCACCAGCTCGA 449
QY 624 CAGCTCAGCTGGCTCGAGTGTACATCGCCCGCGCGCGCGCGCGCGCGCGCGCG 672
Db 450 GAGCGTGGCTTGGCGTGTACAACTACATCTGGGCTTGGCGCGAGACACTG 498

RESULT 8
US-09-042-353-224
; Sequence 224, Application US/09042353
; Patent No. 6255458
; GENERAL INFORMATION:
; APPLICANT: Lonberg, Nils
; APPLICANT: Kay, Robert M.
; TITLE OF INVENTION: Transgenic No. 6255458-Human Animals for
; TITLE OF INVENTION: Producing Heterologous Antibodies
; NUMBER OF SEQUENCES: 421
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/042,353
; FILING DATE: 13-Mar-1998
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/810,279
; FILING DATE: 17-DEC-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/853,408
; FILING DATE: 18-Mar-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/904,068
; FILING DATE: 23-JUN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/990,860
; FILING DATE: 16-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/053,131
; FILING DATE: 26-APR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/096,762
; FILING DATE: 22-JUL-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/155,301
; FILING DATE: 18-NOV-1993
```

```

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/161,739
; FILING DATE: 03-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/165,699
; FILING DATE: 10-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/209,741
; FILING DATE: 09-MAR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/352,322
; FILING DATE: 07-DEC-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/544,404
; FILING DATE: 10-OCT-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/728,463
; FILING DATE: 10-OCT-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/352,322
; FILING DATE: 07-DEC-1994
; APPLICATION NUMBER: WO PCT/US96/16433
; FILING DATE: 10-OCT-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/758,417
; FILING DATE: 02-DEC-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/US97/21803
; FILING DATE: 01-DEC-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Apple, Randolph T.
; REGISTRATION NUMBER: 36,429
; REFERENCE/DOCKET NUMBER: 014643-009040US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 224:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3618 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
; US-09-042-353-224

Query Match 5.0%; Score 63.2; DB 3; Length 3618;
Best Local Similarity 67.4%; Pred. No. 9e-06;
Matches 89; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

QY 144 TGGCTGCGCCACACTCGGAGGCGCTTGGTTTCAGGGTCTCTCTCTCTCTCTCTCTCTCACCC 203
DB 152 TGTGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTC 211
QY 204 TCTTCCTCGGTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTC 263
DB 212 TGTCTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTC 271
QY 264 AACATGTCCAC 275
DB 272 ACACACACAC 283

RESULT 9
US-08-758-417A-72
; Sequence 72, Application US/08758417A
; Patent No. 6300129
; GENERAL INFORMATION:
; APPLICANT: Lonberg, Nils
; Kay, Robert M.
; TITLE OF INVENTION: Transgenic No. 6300129-Human Animals for
; Producing Heterologous Antibodies
; NUMBER OF SEQUENCES: 417
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
;

; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/758,417A
; FILING DATE: 02-DEC-1996
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/728,463
; FILING DATE: 10-OCT-1996
; APPLICATION NUMBER: US 08/544,404
; FILING DATE: 10-OCT-1995
; APPLICATION NUMBER: US 08/352,322
; FILING DATE: 07-DEC-1994
; APPLICATION NUMBER: US 08/209,741
; FILING DATE: 09-MAR-1994
; APPLICATION NUMBER: US 08/165,699
; FILING DATE: 10-DEC-1993
; APPLICATION NUMBER: US 08/161,739
; FILING DATE: 03-DEC-1993
; APPLICATION NUMBER: US 08/155,301
; FILING DATE: 18-NOV-1993
; APPLICATION NUMBER: US 08/096,762
; FILING DATE: 22-JUL-1993
; APPLICATION NUMBER: US 08/053,131
; FILING DATE: 26-APR-1993
; APPLICATION NUMBER: US 07/990,860
; FILING DATE: 16-DEC-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Serafini, Andrew T.
; REGISTRATION NUMBER: 41,303
; REFERENCE/DOCKET NUMBER: 014643-009030US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 72:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3618 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
; FEATURE:
; NAME/KEY: -
; LOCATION: 1..3618
; OTHER INFORMATION: /note= "vector pGPe"
; SEQUENCE DESCRIPTION: SEQ ID NO: 72:
; US-08-758-417A-72

Query Match 5.0%; Score 63.2; DB 3; Length 3618;
Best Local Similarity 67.4%; Pred. No. 9e-06;
Matches 89; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

QY 144 TGGCTGCGCCACACTCGGAGGCGCTTGGTTTCAGGGTCTCTCTCTCTCTCTCTCTCTCACCC 203
DB 152 TGTGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTC 211
QY 204 TCTTCCTCGGTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTC 263
DB 212 TGTCTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTC 271
QY 264 AACATGTCCAC 275
DB 272 ACACACACAC 283

RESULT 10
```

[illegible]

REGISTRATION NUMBER: 87654  
REFERENCE/DOCKET NUMBER: 14643-5  
TELEPHONE: 415-543-9600  
TELEFAX: 415-543-5043  
INFORMATION FOR SEQ ID NO: 49:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3698 base pairs  
TYPE: NUCLEIC ACID  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
PCT-US92-06185-49

Query Match  
Best Local Similarity 5.0%; Score 63.2; DB 5; Length 3698;  
Matches 89; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

Qy 144 TGGCTGGCCACACTCGGAGGCGCTTGGTTTCAGGGTCTCTGTCCTCTCTCACC 203  
Db 153 TGTGTGTGTC 212  
Qy 204 TCTTCCTCGCTTCTCTGTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTC 263  
Db 213 TGTCTGTC 272  
Qy 264 AACATGTCCAC 275  
Db 273 ACACACACAC 284

RESULT 13  
US-08-053-131-120  
Sequence 120, Application US/08053131  
Patent No. 5661016  
GENERAL INFORMATION:  
APPLICANT: Lonberg, Nils  
APPLICANT: Kay, Robert M.  
TITLE OF INVENTION: Transgenic No. 5661016-Human Animals for  
TITLE OF INVENTION: Producing Heterologous Antibodies  
NUMBER OF SEQUENCES: 197  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend Kourie and Crew  
STREET: One Market Plaza, Steuart Tower, Suite 200  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94105  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/053,131  
FILING DATE: 26-APR-1993  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/990,860  
FILING DATE: 16-DEC-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/810,279  
FILING DATE: 17-DEC-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/853,408  
FILING DATE: 18-MAR-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Smith, William M.  
REGISTRATION NUMBER: 30,223  
REFERENCE/DOCKET NUMBER: 14643-9-3  
TELEPHONE: 415-326-2400  
TELEFAX: 415-326-2422

INFORMATION FOR SEQ ID NO: 120:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3699 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-053-131-120

Query Match  
Best Local Similarity 5.0%; Score 63.2; DB 1; Length 3699;  
Matches 89; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

Qy 144 TGGCTGGCCACACTCGGAGGCGCTTGGTTTCAGGGTCTCTGTCCTCTCTCACC 203  
Db 153 TGTGTGTGTC 212  
Qy 204 TCTTCCTCGCTTCTCTGTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTC 263  
Db 213 TGTCTGTC 272  
Qy 264 AACATGTCCAC 275  
Db 273 ACACACACAC 284

RESULT 14  
US-08-645-641-120  
Sequence 120, Application US/08645641  
Patent No. 5719032  
GENERAL INFORMATION:  
APPLICANT: Lonberg, Nils  
APPLICANT: Kay, Robert M.  
TITLE OF INVENTION: Transgenic No. 5719032-Human Animals for  
TITLE OF INVENTION: Producing Heterologous Antibodies  
NUMBER OF SEQUENCES: 150  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: William M. Smith  
STREET: Two Embarcadero Center, 8th Floor  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-3834  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/645,641  
FILING DATE: 20-MAY-1996  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/904,068  
FILING DATE: 23-JUN-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Smith, William M.  
REGISTRATION NUMBER: 30,223  
REFERENCE/DOCKET NUMBER: 14643-000913  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-326-2400  
TELEFAX: 415-326-2422  
INFORMATION FOR SEQ ID NO: 120:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3699 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-645-641-120

Query Match  
Best Local Similarity 5.0%; Score 63.2; DB 1; Length 3699;  
Matches 89; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

Matches 89; Conservative 0; Mismatches 43; Indels 0; Gaps 0;  
QY 144 TGGCTCGGCACACTCGGAGGCTTGGTTTCAGGGTCTCTGTCTCTCTCTCACCC 203  
Db 153 TGGTGTGTC 212  
QY 204 TCTTCCTCGCTTC 263  
Db 213 TGTCTCTGTC 272  
QY 264 AACATGTCCAC 275  
Db 273 ACACACACAC 284

## RESULT 15

US-07-853-408B-120  
; Sequence 120, Application US/07853408B  
; Patent No. 5789650  
; GENERAL INFORMATION:  
; APPLICANT: Lonberg, Nils  
; APPLICANT: Kay, Robert M.  
; TITLE OF INVENTION: Transgenic No. 5789650-Human Animals for  
; TITLE OF INVENTION: Producing Heterologous Antibodies  
; NUMBER OF SEQUENCES: 150  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: William M. Smith  
; STREET: One Market Plaza, Steuart Tower, Suite 2000  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94105  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/853,408B  
; FILING DATE: 19920318  
; CLASSIFICATION: 800  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Smith, William M.  
; REGISTRATION NUMBER: 30,223  
; REFERENCE/DOCKET NUMBER: 14643-9  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-326-2400  
; TELEFAX: 415-326-2422  
; INFORMATION FOR SEQ ID NO: 120:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3699 base pairs  
; TYPE: NUCLEIC ACID  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
US-07-853-408B-120

Query Match 5.0%; Score 63.2; DB 1; Length 3699;  
Best Local Similarity 67.4%; Pred. No. 9.1e-06;  
Matches 89; Conservative 0; Mismatches 43; Indels 0; Gaps 0;  
QY 144 TGGCTCGGCACACTCGGAGGCTTGGTTTCAGGGTCTCTGTCTCTCTCTCACCC 203  
Db 153 TGGTGTGTC 212  
QY 204 TCTTCCTCGCTTC 263  
Db 213 TGTCTCTGTC 272  
QY 264 AACATGTCCAC 275  
Db 273 ACACACACAC 284

GenCore version 5.1.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: December 17, 2004, 00:04:02 ; Search time 455 Seconds

(without alignments)  
2065.157 Million cell updates/sec

Title: US-09-701-674A-23

Perfect score: 179

Sequence: 1 MSTGSLSDVDLQVEVMELE.....PESDLKEYVTASRLCGTTAS 179

Scoring table:

OLIGO 60.0 , Xgapext 60.0  
Ygapop 60.0 , Ygapext 60.0  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 4134886 seqs, 2624710521 residues

Word size: 1

Total number of hits satisfying chosen parameters: 8266295

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters: -DB=xlh

-MODEL=frame\_plus\_p2n.model -DEV=xlh  
-Q=/cgn2.1/USPTO spool\_h/US03701674/runat\_15122004\_100529\_28299/app\_query.fasta\_1.327  
-DB=N Geneseq 23Sep04 -QFMT=fastap -SUFFIX=oligo.rng -MINMATCH=0.1 -LOOPEL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi  
-LIST=45 -DOCALLGN=200 -THR SCORE=quality -THR MIN=1 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US03701674 @CGN\_1\_1\_470 @runat\_15122004\_100529\_28299 -NCFU=6 -ICPU=3  
-NO\_WMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : N\_Geneseq\_23Sep04:\*

1: Geneseqn1980s: \*  
2: Geneseqn1990s: \*  
3: Geneseqn2000s: \*  
4: Geneseqn2001as: \*  
5: Geneseqn2001bs: \*  
6: Geneseqn2002as: \*  
7: Geneseqn2002bs: \*  
8: Geneseqn2003as: \*  
9: Geneseqn2003bs: \*  
10: Geneseqn2003cs: \*  
11: Geneseqn2003ds: \*  
12: Geneseqn2004s: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	179	100.0	1272	3	Aaz57861 Protein r
2	94	52.5	482	3	Ach21857 Human adu
3	61	34.1	697	6	Ab165519 Lung canc
4	61	34.1	697	6	Ab165935 Lung canc
5	61	34.1	697	6	Abk64784 Human ben
6	49	27.4	632	6	Abq55427 Human ova

7	36	20.1	240	10	ACA56430	Acas6430 Chicken s
8	36	20.1	240	12	ADJ56226	Adj56226 Human pol
9	28	15.6	493	9	ACH23369	Ach23369 Human adu
10	28	15.6	1716	3	AAZ50465	Aaz50465 Human mus
11	28	15.6	1718	3	AAZ50464	Aaz50464 Murine mu
12	28	15.6	2177	10	ADZ52593	Adz52593 Human cDN
13	28	15.6	2196	3	AAA72428	Aaa72428 Human nuc
14	28	15.6	2196	10	ADJ56314	Adj56314 Human cDN
15	28	15.6	2382	6	AAS94828	Aas94828 Human DNA
16	18	10.1	85	6	ABN27386	Abn27386 Rat splc
17	15	8.4	1086	4	ABL08723	Ab108723 Drosophil
18	15	8.4	5101	4	ABL08722	Ab108722 Drosophil
19	12	6.7	963	8	ABZ18589	Abz18589 Group III
20	11	6.1	474	5	AAS76438	Aas76438 DNA encod
21	11	6.1	966	5	AAS76439	Aas76439 DNA encod
22	9	5.0	730	4	AAS12517	Aas12517 Gene #10
23	9	5.0	730	10	ABQ77254	Abq77254 Human neu
24	9	5.0	981	10	ABQ51583	Abq51583 Human str
25	9	5.0	1140	6	ABQ88217	Abq88217 Human ost
26	9	5.0	2610	11	ACH95595	Ach95595 Klebsiell
27	9	5.0	2787	4	ABL10299	Ab110299 Drosophil
28	9	5.0	4852	4	ABL10298	Ab110298 Drosophil
29	9	5.0	6288	5	AAS45090	Aas45090 cDNA encod
30	9	5.0	12905	3	AAS99101	Aaz99101 S. fradia
31	9	5.0	12905	3	AAZ99100	Aaz99100 S. fradia
32	8	4.5	276	12	ADQ07882	Adq07882 Soybean e
33	8	4.5	302	8	ABX55437	Abx55437 Bovine ES
34	8	4.5	353	10	ADK57797	Adk57797 Plant DNA
35	8	4.5	375	10	AB241346	Ab241346 N. gonorr
36	8	4.5	375	10	AB240186	Ab240186 N. gonorr
37	8	4.5	378	3	AAA75746	Aaa75746 DNA encod
38	8	4.5	381	8	ABX09886	Abx09886 N. mening
39	8	4.5	405	8	ABX55489	Abx55489 Bovine ES
40	8	4.5	420	3	AAZ54169	Aaz54169 Neisseria
41	8	4.5	420	3	AAZ54168	Aaz54168 Neisseria
42	8	4.5	420	3	AAZ54170	Aaz54170 Neisseria
43	8	4.5	423	10	AB240195	Ab240195 N. gonorr
44	8	4.5	423	10	AB241348	Ab241348 N. gonorr
45	8	4.5	518	6	ABQ52056	Abq52056 Oligonuc

#### ALIGNMENTS

RESULT 1

AZ57861

ID AZ57861 standard; cDNA; 1272 BP.

XX AC

XX AC AZ57861;

XX DT

DT 11-APR-2000 (first entry)

XX DE

DE Protein regulating gene expression PRGE-23 cDNA clone 3340296.

XX XX

XX Protein regulating gene expression; PRGE-23; human; cancer; inflammation;  
XX anticancer; antitumour; antiinflammatory; Myc; HLH protein; gene therapy;  
XX diagnosis; ss.

XX OS

OS Homo sapiens.

XX XX

XX Key Location/Qualifiers

XX FT 268..807

XX FT /\*tag= a

XX XX

XX WO9964596-A2.

XX XX

XX 16-DEC-1999.

XX XX

XX 11-JUN-1999; 99WO-US013281.

XX XX

XX 12-JUN-1998; 98US-0089029P.

XX PR

PR 29-JUL-1998; 98US-0094575P.

XX PR

PR 14-OCT-1998; 98US-0104624P.

XX XX



PA (INCY-) INCYTE PHARM INC.  
 XX Lail P, Yue H, Tang YT, Hillman JL, Bandman O, Corley NC;  
 PI Guegler KJ, Gorgone GA, Baughn MR, Patterson C, Lu DM;  
 XX  
 DR WPI; 2000-116543/10.  
 DR P-PSDB; AAY58630.  
 XX  
 XX New human polypeptides that regulate gene expression, for treatment,  
 PT prevention and diagnosis of, e.g. cancer.  
 XX  
 XX Claim 9; Page 139; 150pp; English.  
 XX  
 CC The present sequence is that of Incyte clone 3340296 encoding new human  
 CC protein regulating gene expression PRGE-23 (see AAY58630). The cDNA was  
 CC initially isolated from spleen tissue cDNA library SPLNNOT10, and the  
 CC full-length sequence assembled from overlapping clones from a number of  
 CC libraries. PRGE-23 is expressed in reproductive, developmental and  
 CC urologic tissues associated with cancer, inflammation and foetal  
 CC diseases, disorders or conditions. It is characterised as an Myc-type HLH  
 CC protein. The invention provides PRGE polypeptides (see AAY58608-38) and  
 CC polynucleotides (see AA257839-69), expression vectors, host cells,  
 CC antibodies, agonists and antagonists. It also provides methods for  
 CC diagnosing, treating or preventing disorders associated with expression  
 CC of PRGE. Polynucleotides are also used as sources of probes and primers  
 CC for diagnosis and monitoring of disease, also for detecting related  
 CC sequences and in gene mapping  
 XX  
 SQ Sequence 1272 BP; 311 A; 358 C; 318 G; 285 T; 0 U; 0 Other;  
 Alignment Scores:  
 Pred. No.: 2,118-172 Length: 1272  
 Score: 179.00 Matches: 179  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 3 Gaps: 0  
 US-09-701-674A-23 (1-179) x AA257861 (1-1272)  
 QY 1 MetSerThrGlySerIleuSerAspValGluAspLeuGlnGluValGluMetLeuGluCys 20  
 DB 268 ATGTCACCGCGTCCCTCAGCAGTGTGGAGGACCTTCAAGAGGTGGAGATGTGGAATGT 327  
 QY 21 AspGlyLeuIleuMetAspSerAsnIleuValThrSerAsnGluSerThrGluGlu 40  
 DB 328 GACGGGTTCAAATGATTCGAACCAAGGAATTTGTGACTTCAACGAGCAGCAGGAG 387  
 QY 41 SerSerAsnGlyGluAsnGlySerProGlnIleuValThrSerAsnGlyLeuGlyLeuArg 60  
 DB 388 AGCTCCAACTCGAGAAATGGGTCTCCCAAGAGGCGCGCGGCTCGGCAAGAGGAGG 447  
 QY 61 LysAlaProThrIleuSerProIleuSerGlyValSerGlnGluGlyValGlnValGln 80  
 DB 448 AAGGGCGCCCAACAGAGAGCCCTGAGCGGGGTGACCGAGGAGGAGGAGGAGGAGGAG 507  
 QY 81 ArgAsnAlaAlaAsnAlaArgGluArgAlaArgMetArgValLeuSerIleuAlaPheSer 100  
 DB 508 CGCAAGCGCGCCCAACAGAGGAGGCGGCGCCGATCGAGTGTCTGAGCAGGCTTCTCC 567  
 QY 101 ArgLeuIleuThrIleuProThrValProProAspThrLysLeuSerIleuAspThr 120  
 DB 568 AGACTCAAGACCACTGCTGCGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 627  
 QY 121 LeuArgLeuAlaSerSerIleuAlaHisLeuArgGlnIleuLeuAlaAsnAspIleu 140  
 DB 628 CTCAGCTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 687  
 QY 141 GluAsnGlyThrIleHisProValAsnLeuThrIleProPheMetValAlaGlyLysPro 160  
 DB 688 GAGAACGGGTGATTCATTCACCGGTCAACCTGACGTGGCGCTTTATGGTGGCGGGAACCC 747  
 QY 161 GluSerAspLeuIleuValThrAlaSerArgLeuGlyThrAlaSer 179

DB 748 GAGAGTGACCTGAAGAGAGTGTGACCGGAGCGCTTATGTGGAACCAACCGCGTCC 804

RESULT 2  
 ACH21857  
 ID ACH21857 standard; cDNA; 482 BP.  
 XX AC ACH21857;  
 XX DT 13-OCT-2003 (first entry)  
 XX DE Human adult ovary cDNA #237.  
 XX KW Human; ss; sequencing by hybridisation; SBH; expressed sequence tag; EST;  
 XX KM genome mapping; biodiversity; genetic disorder.  
 XX OS Homo sapiens.  
 XX PN US2003073623-A1.  
 XX PD 17-APR-2003.  
 XX PF 30-JUL-2001; 2001US-00918995.  
 XX PR 30-JUL-2001; 2001US-00918995.  
 XX PA (DRMA/) DRMANAC R T.  
 PA (LABA/) LABAT I.  
 PA (STAC/) STACHE-CRAIN B.  
 PA (DICK/) DICKSON M C.  
 PA (JONE/) JONES L W.  
 XX PI Drmanac RT, Labat I, Stache-Crain B, Dickson MC, Jones LW;  
 XX WPI; 2003-615964/58.  
 XX New polynucleotide sequences obtained from various cDNA libraries, useful  
 PT as hybridization probes, as oligomers for PCR, for chromosome and gene  
 PT mapping, in the recombinant production of protein, or in generating  
 PT antisense DNA or RNA.  
 XX Claim 1; SEQ ID NO 9069; 44pp; English.  
 CC The invention relates to an isolated polynucleotide comprising any one of  
 CC 38043 cDNA sequences, appearing as ACH21857-ACH50831, whose sequence was  
 CC determined by the technique of SBH (sequencing by hybridisation). Also  
 CC included is a purified polypeptide comprising a sequence corresponding to  
 CC a reading frame of the novel polynucleotide. The nucleic acid sequences  
 CC are useful in diagnostics as expressed sequence tags (EST) for  
 CC identifying expressed genes or for physical mapping of the human genome,  
 CC in forensics, in assessing biodiversity, or in identifying mutations  
 CC responsible for genetic disorders and other traits. The nucleotide  
 CC sequences are also useful as hybridisation probes, as oligomers for PCR,  
 CC for chromosome and gene mapping, in the recombinant production of  
 CC protein, or in generating antisense DNA or RNA. The purified polypeptide  
 CC is useful for generating antibodies specific for it. The present sequence  
 CC is one of the 38043 isolated cDNA/EST sequences. Note: The sequence data  
 CC for this patent did not form part of the printed specification, but was  
 CC obtained in electronic format directly from USPTO at  
 CC seqdata.uspto.gov/sequence.html?DocID=20030073623  
 XX SQ Sequence 482 BP; 99 A; 156 C; 143 G; 83 T; 0 U; 1 Other;  
 Alignment Scores:  
 Pred. No.: 3,458-86 Length: 482  
 Score: 94.00 Matches: 94  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 52.51% Indels: 0  
 DB: 9 Gaps: 0  
 US-09-701-674A-23 (1-179) x ACH21857 (1-482)

QY 86 AlaArgGluArgAlaArgVetArgValLeuSerLysAlaPheSerArgLeuLysThrThr 105  
 Db 39 GCGGAGAGCGGCGCGATGGAGTGTCTGAGCAAGCGCTTCTCCAGACTCAAGACCCACC 98  
 QY 106 LeuProTirValProProAspThrLysLeuSerLysLeuAspThrLeuArgLeuAlaSer 125  
 Db 99 CTGCGCTGGTGGCGCCCGACACCAAGCTCTCAGCTGGACACGCTCAGGCTGGCGTCC 158  
 QY 126 SerTyrIleAlaHisLeuArgGlnIleLeuAlaAsnAspLysTyrGluAsnGlyTyrIle 145  
 Db 159 AGTACATCGCCCACTTGAGGCGAGATCTCTGGCTAACGACAAATACGAGAACGGGTACATT 218  
 QY 145 HisProValAsnLeuThrTrpProPheMetValAlaGlyLysProGluSerAspLeuLys 165  
 Db 219 CACCGCGTCAACTGACGTGGCGCCCTTATGGTGGCGCGGAAACCCGAGAGTGACCTGAAA 278  
 QY 166 GluValValThrAlaSerArgLeuCysGlyThrThrAlaSer 179  
 Db 279 GAAGTGGTACCGCGAGCGCGCTTATGTGGAACCCCGCGTCC 320

## RESULT 3

ABL65519

ID ABL65519 standard; DNA; 697 BP.

XX ABL65519;

AC

XX

DT 15-MAY-2002 (first entry)

XX

DE Lung cancer related gene sequence SEQ ID NO:3856.

XX

KW Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;

KW stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;

KW cytostatic; gene therapy; antineoplastic; Wilm's tumour; adenocarcinoma;

KW gene; db.

XX Homo sapiens.

OS

XX WO200194629-A2.

PN

XX

PD 13-DEC-2001.

XX

PF 30-MAY-2001; 2001WO-US010838.

XX

PR 05-JUN-2000; 2000US-0209473P.

PR 05-JUN-2000; 2000US-0209531P.

PR 18-SEP-2000; 2000US-0233133P.

PR 18-SEP-2000; 2000US-0233161P.

PR 20-SEP-2000; 2000US-0234009P.

PR 20-SEP-2000; 2000US-0234034P.

PR 20-SEP-2000; 2000US-0234052P.

PR 22-SEP-2000; 2000US-0234509P.

PR 22-SEP-2000; 2000US-0234567P.

PR 25-SEP-2000; 2000US-0234923P.

PR 25-SEP-2000; 2000US-0234924P.

PR 25-SEP-2000; 2000US-0235077P.

PR 25-SEP-2000; 2000US-0235082P.

PR 25-SEP-2000; 2000US-0235134P.

PR 26-SEP-2000; 2000US-0235280P.

PR 26-SEP-2000; 2000US-0235637P.

PR 26-SEP-2000; 2000US-0235638P.

PR 27-SEP-2000; 2000US-0235711P.

PR 27-SEP-2000; 2000US-0235720P.

PR 27-SEP-2000; 2000US-0235840P.

PR 27-SEP-2000; 2000US-0235863P.

PR 28-SEP-2000; 2000US-0236028P.

PR 28-SEP-2000; 2000US-0236032P.

PR 28-SEP-2000; 2000US-0236033P.

PR 28-SEP-2000; 2000US-0236034P.

PR 28-SEP-2000; 2000US-0236109P.

PR 28-SEP-2000; 2000US-0236111P.

PR 29-SEP-2000; 2000US-0236842P.

PR 29-SEP-2000; 2000US-0236891P.

PR 02-OCT-2000; 2000US-0237172P.

PR 02-OCT-2000; 2000US-0237173P.  
 PR 02-OCT-2000; 2000US-0237278P.  
 PR 02-OCT-2000; 2000US-0237294P.  
 PR 02-OCT-2000; 2000US-0237295P.  
 PR 02-OCT-2000; 2000US-0237316P.  
 PR 03-OCT-2000; 2000US-0237425P.  
 PR 03-OCT-2000; 2000US-0237598P.  
 PR 03-OCT-2000; 2000US-0237604P.  
 PR 03-OCT-2000; 2000US-0237606P.  
 PR 03-OCT-2000; 2000US-0237608P.  
 PR 01-NOV-2000; 2000US-0244867P.  
 PR 01-NOV-2000; 2000US-0245084P.

XX (AVAL-) AVALON PHARM.

XX Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;

PI Soppet DR, Weaver Z;

XX WPI; 2002-188264/24.

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

Screening for anti-neoplastic agent involves exposing cells to a chemical agent to be tested for anti-neoplastic activity, and determining a change in expression of a gene of a signature gene set.

Claim 1; SEQ ID NO 3856; 44pp; English.

The present invention describes a method (M1) for screening for an anti-neoplastic agent. The method involves exposing cells to a chemical agent to be tested for anti-neoplastic activity, determining a change in expression of at least one gene (I) of a signature gene set, where (I) comprises a sequence (S) selected from 8447 sequences (given in ABL61664 to ABL70110), or is at least 95% identical to (S), where a change in expression is indicative of anti-neoplastic activity. (I) has cytostatic activity and can be used in gene therapy. M1 can be used for screening an anti-neoplastic agent, and can be used for producing a product which is the data collected with respect to the anti-neoplastic agent as a result of M1, and the data is sufficient to convey the chemical structure and/or properties of the agent. M1 can be used in the treatment of cancer such as colon, breast, stomach, lung, thyroid, oesophageal, ovarian, kidney, prostate or pancreatic cancer, adenocarcinoma, carcinoma, clear cell cancer, infiltrating ductal cancer, infiltrating lobular cancer, squamous cell carcinoma, neuroendocrine carcinoma, papillary carcinoma and Wilm's tumour

Sequence 697 BP; 190 A; 164 C; 156 G; 180 T; 0 U; 7 Other;

## Alignment Scores:

Pred. No.: 2.06e-52 Length: 697  
 Score: 61.00 Matches: 61  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 34.08% Indels: 0  
 DB: 6 Gaps: 0

US-09-701-674A-23 (1-179) x ABL65519 (1-697)

Qy 119 AspThrLeuArgLeuAlaSerSerTyrIleAlaHisLeuArgGlnIleLeuAlaAsnAsp 138

Db 3 GACACGCTCAGGCTGGCGTCCAGCTACATCGCCACTTGAGGCGAGATCCTGGCTAACGAC 62

Qy 139 LysTyrGluAsnGlyTyrIleHisProValAsnLeuThrTrpProPheMetValAlaGly 158

Db 63 AAATACGAGAACGGGTACATTACCCGGTCAACCTGAGCGCCCTTTATGGTGGCGGG 122

Qy 159 LysProGluSerAspLeuLysGluValValThrAlaSerArgLeuCysGlyThrThrAla 178

Db 123 AAACCCGAGAGTGCCTGAAAGAGTGGTACCCGCGAGCGCCTTATGTGGAACACCGCG 182

Qy 179 Ser 179

Db 183 TCC 185

RESULT 4

```

ABL65935
ID ABL65935 standard; DNA; 697 BP.
XX
AC ABL65935;
XX
DT 15-MAY-2002 (first entry)
XX
DE Lung cancer related gene sequence SEQ ID NO:4272.
XX
KW Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;
KW stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;
KW cytostatic; gene therapy; antineoplastic; Wilm's tumour; adenocarcinoma;
KW gene; ds.
XX
OS Homo sapiens.
XX
PN WO200194629-A2.
PD
XX
PD 13-DEC-2001.
XX
XX
XX 30-MAY-2001; 2001WO-US010838.
XX
PR 05-JUN-2000; 2000US-0209473P.
PR 05-JUN-2000; 2000US-0209531P.
PR 18-SEP-2000; 2000US-0233133P.
PR 18-SEP-2000; 2000US-0233617P.
PR 20-SEP-2000; 2000US-0234009P.
PR 20-SEP-2000; 2000US-0234034P.
PR 20-SEP-2000; 2000US-0234052P.
PR 22-SEP-2000; 2000US-0234509P.
PR 22-SEP-2000; 2000US-0234567P.
PR 25-SEP-2000; 2000US-0234923P.
PR 25-SEP-2000; 2000US-0234924P.
PR 25-SEP-2000; 2000US-0235077P.
PR 25-SEP-2000; 2000US-0235082P.
PR 25-SEP-2000; 2000US-0235134P.
PR 25-SEP-2000; 2000US-0235280P.
PR 26-SEP-2000; 2000US-0235637P.
PR 26-SEP-2000; 2000US-0235638P.
PR 27-SEP-2000; 2000US-0235711P.
PR 27-SEP-2000; 2000US-0235720P.
PR 27-SEP-2000; 2000US-0235840P.
PR 27-SEP-2000; 2000US-0235863P.
PR 28-SEP-2000; 2000US-0236028P.
PR 28-SEP-2000; 2000US-0236032P.
PR 28-SEP-2000; 2000US-0236033P.
PR 28-SEP-2000; 2000US-0236034P.
PR 28-SEP-2000; 2000US-0236109P.
PR 28-SEP-2000; 2000US-0236111P.
PR 29-SEP-2000; 2000US-0236842P.
PR 29-SEP-2000; 2000US-0236891P.
PR 02-OCT-2000; 2000US-0237172P.
PR 02-OCT-2000; 2000US-0237173P.
PR 02-OCT-2000; 2000US-0237278P.
PR 02-OCT-2000; 2000US-0237294P.
PR 02-OCT-2000; 2000US-0237295P.
PR 02-OCT-2000; 2000US-0237316P.
PR 03-OCT-2000; 2000US-0237425P.
PR 03-OCT-2000; 2000US-0237598P.
PR 03-OCT-2000; 2000US-0237604P.
PR 03-OCT-2000; 2000US-0237606P.
PR 03-OCT-2000; 2000US-0237608P.
PR 01-NOV-2000; 2000US-0244867P.
PR 01-NOV-2000; 2000US-0245084P.
XX
XX (AVAL-) AVALON PHARM.
XX
XX Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;
XX Soppet DR, Weaver Z;
XX WPI; 2002-188264/24.
XX
XX Screening for anti-neoplastic agent involves exposing cells to a chemical
PT agent to be tested for anti-neoplastic activity, and determining a change
PT in expression of a gene of a signature gene set.
XX
XX Claim 1; SEQ ID NO 4272; 4pp; English.
XX
XX The present invention describes a method (M1) for screening for an anti-
XX neoplastic agent. The method involves exposing cells to a chemical agent
XX to be tested for anti-neoplastic activity, determining a change in
XX expression of at least one gene (I) of a signature gene set, where (I)
XX comprises a sequence (S) selected from 8447 sequences (given in ABL61664
XX to ABL70110), or is at least 95% identical to (S), where a change in
XX expression is indicative of anti-neoplastic activity. (I) has cytosolic
XX activity and can be used in gene therapy. M1 can be used for screening an
XX anti-neoplastic agent, and can be used for producing a product which is
XX the data collected with respect to the anti-neoplastic agent as a result
XX of M1, and the data is sufficient to convey the chemical structure and/or
XX properties of the agent. M1 can be used in the treatment of cancer such
XX as colon, breast, stomach, lung, thyroid, oesophageal, ovarian, kidney,
XX prostate or pancreatic cancer, adenocarcinoma, carcinoma, clear cell
XX cancer, infiltrating ductal cancer, infiltrating lobular cancer, squamous
XX cell carcinoma, neuroendocrine carcinoma, papillary carcinoma and Wilm's
XX tumour
XX
XX Sequence 697 BP; 190 A; 164 C; 156 G; 180 T; 0 U; 7 Other;
SQ
Alignment Scores:
Pred. No.: 2.06e-52 Length: 697
Score: 61.00 Matches: 61
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 34.08% Indels: 0
DB: 6 Gaps: 0
US-09-701-674A-23 (1-179) x ABL65935 (1-697)
Qy 119 AspThrLeuArgLeuAlaSerSerTyrIleAlaHisLeuArgGlnrleuAlaAsnAsp 138
Db 3 GACACGCTCAGGCTGGGTCAGCTACATCGCCACTTCAGGCAGATCTCTGGCTAACGAC 62
Qy 139 LysTyrGluAsnGlyTyrIleHisProValAsnLeuThrTrpProPheMetValAlaGly 158
Db 63 AAATACGAGAACGGGTACATTACCCCGGTCAACCTGACGTGCGCCCTTTATGTGGCCGG 122
Qy 159 LysProGluSerAspLeuLysGluValValThrAlaSerArgLeuCysGlyThrThrAla 178
Db 123 AAACCCGAGAGTGCCTGAAAGAGTGTGACCGCGCGCTTATGTGACACCCGCG 182
Qy 179 Ser 179
Db 183 TCC 185
RESULT 5
ABK64784
ID ABK64784 standard; DNA; 697 BP.
XX
XX ABK64784;
XX
XX AC ABK64784;
XX
XX DT 18-JUN-2002 (first entry)
XX
XX Human benign prostatic hyperplasia gene #679.
XX
XX Human; benign prostatic hyperplasia; BPH; prostate cancer; gene; ds.
XX
XX Homo sapiens.
XX
XX WO200212440-A2.
XX
XX 14-FEB-2002.
XX
XX 07-AUG-2001; 2001WO-US024708.
XX
XX 07-AUG-2000; 2000US-0223323P.
XX
XX 05-JUN-2001; 2001US-00873319.

```

XX (GENE-) GENE LOGIC INC.  
PA (NLSB ) JAPAN TOBACCO INC.  
XX Munger WE, Kulkarni P, Getzenberg RH, Waga I, Yamamoto J;  
XX WPI; 2002-257476/30.  
XX Identifying drugs for and diagnosing benign prostatic hyperplasia, by  
PT detecting expression levels of one or more genes in prostate cells from  
PT patient that are differentially regulated compared to normal prostate  
PT cells.  
XX Disclosure; Page 367; 443pp; English.  
XX  
XX The invention relates to a method of diagnosing (I) the onset or  
CC progression of benign prostatic hyperplasia (BPH), or screening (II) for  
CC or identifying an agent that modulates the onset or progression of BPH.  
CC The method is based on changes in gene expression in BPH tissue isolated  
CC from patients exhibiting different clinical states of prostate  
CC hyperplasia as compared to normal prostate tissue. (I) comprises  
CC detecting the expression levels of one or more genes in prostate cells  
CC from the subject that are differentially regulated compared to normal  
CC prostate cells. (II) comprises preparing a first gene expression profile  
CC of BPH cells or BPH-like cell population, exposing the cells to the  
CC agent, preparing a second gene expression profile of the agent exposed  
CC cells, and comparing the first and second gene expression profiles. (I)  
CC is useful for diagnosing the onset or progression of BPH. (II) is useful  
CC for identifying an agent that modulates the onset or progression of BPH.  
CC The methods are useful to present information identifying the expression  
CC level in a tissue or cells, by comparing the expression level of genes  
CC given in the specification in the tissue or cells to the level of  
CC expression of gene in the database, and displaying the expression levels  
CC of at least one gene in the tissue or cell sample compared to the  
CC expression level in BPH. Agents using (II) are useful for treating BPH or  
CC prostate cancer. ABK4106-ABK4860 represent human benign prostatic  
CC hyperplasia gene sequences of the invention  
XX  
XX Sequence 697 BP; 190 A; 164 C; 156 G; 180 T; 0 U; 7 Other;  
SQ

Alignment Scores:  
Pred. No.: 2,066-52 Length: 697  
Score: 61.00 Matches: 61  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 34.08% Indels: 0  
DB: 6 Gaps: 0

US-09-701-674A-23 (1-179) x ABK64784 (1-697)

QY 119 AspThrLeuArgLeuAlaSerSerrTrilleAlaHisLeuArgGlnIleLeuAlaHisAsp 138  
Db 3 GACACGCTCAGGTGGCGTCCAGTACATCGCCACTTCAGGCAGATCTGGCTAACGAC 62  
QY 139 LysTyvGluAsnGlyTyvTrilleHisProValanLeuThrTrpProPheMetValalagly 158  
Db 63 AAATACGAGAACGGGTACATTACCCGGTCAACCTGACGTGGCCCTTTATGTGTGCCGGG 122  
QY 159 LysProGluSerAspLeuLysGluValValThrAlaSerArgLeuCysGlyThrThrAla 178  
Db 123 AAACCCGAGAGTGACCTGAAGAAGTGTGTACCGGCGCGCGCTTATGTGGACCAACCGCG 182  
QY 179 Ser 179  
Db 183 TCC 185

RESULT 6  
ID ABQ55427 standard; cDNA; 632 BP.  
XX AC ABQ55427;  
XX 22-AUG-2002 (first entry)  
DT

XX Human ovarian antigen HNOJ185 cDNA, SEQ ID NO:1307.  
DE  
XX Human; ovarian antigen; ovary; ovarian; breast; cancer; tumour;  
KW ovarian cancer; breast cancer; tumour; reproductive system disorder;  
KW infertility; pregnancy disorder; anovulation; polycystic ovary syndrome;  
KW PCOS; ovarian cyst; dysmenorrhea; endocrine disorder; infection;  
KW inflammatory condition; immune disorder; blood disorder;  
KW cardiovascular disorder; respiratory disorder; neurological disorder;  
KW gastrointestinal disorder; urinary system disorder; drug screening;  
KW gene therapy; chromosome mapping; forensic analysis;  
KW antibody preparation; cytostatic; immunomodulatory; neuroprotective;  
KW antiinflammatory; gynaecological; reproductive; gene; ss.  
XX  
XX Homo sapiens.  
OS  
XX WO200200677-A1.  
FN  
XX 03-JAN-2002.  
PD  
XX 07-JUN-2001; 2001WO-US018569.  
PF  
XX 07-JUN-2000; 2000US-0209467P.  
PR  
XX (HUMA-) HUMAN GENOME SCI INC.  
FA  
XX Birse CE, Rosen CA;  
PI  
XX WPI; 2002-147878/19.  
DR  
XX P-PSDB; ABP42350.  
DR  
XX Isolated nucleic acid molecules encoding novel ovarian polypeptides,  
PT useful in the prevention, treatment and diagnosis of cancer (e.g. ovarian  
PT cancer), immune disorders, cardiovascular disorders and neurological  
PT diseases.  
XX  
XX Claim 1; SEQ ID NO 1307; 2922pp; English.  
PS  
XX The invention relates to 2175 novel human ovarian antigens (ABP41054-  
CC ABP43228) and to cDNAs encoding them (ABQ54131-ABQ56305), and also  
CC encompasses polypeptides 90% identical and polynucleotides 95% identical  
CC to the sequences of the invention. The invention additionally relates to  
CC recombinant vectors and host cells comprising human ovarian antigen  
CC polynucleotides, antibodies against human ovarian antigens, and the use  
CC of ovarian antigen polynucleotides and polypeptides in diagnosing,  
CC treating, prognosing or preventing various ovary and/or breast-related  
CC disorders. Such conditions include ovarian cancer and breast cancer, and  
CC metastatic tumours of ovarian or breast origin, reproductive system,  
CC disorders (e.g., infertility, disorders of pregnancy, anovulation,  
CC polycystic ovary syndrome, ovarian cysts, and dysmenorrhoea), endocrine  
CC disorders, infections (e.g., chlamydia, HIV, toxoplasmosis, and toxic  
CC shock syndrome), inflammatory conditions (e.g., mastitis, oophoritis and  
CC vaginitis), immune disorders (e.g., congenital and acquired  
CC immunodeficiencies, autoimmune oophoritis, systemic lupus erythematosus),  
CC blood-related disorders (e.g., anaemia), cardiovascular disorders,  
CC respiratory disorders, neurological disorders, gastrointestinal disorders  
CC and urinary system disorders. Ovarian antigen polypeptides and  
CC polynucleotides may also be used in screening for compounds which  
CC modulate ovarian antigen expression or activity. The polynucleotides may  
CC further be used for gene therapy, chromosome mapping, in the  
CC identification of individuals and in forensic analysis, and the  
CC polypeptides may be used as food additives or to prepare antibodies  
CC useful in disease diagnosis, drug targeting and phenotyping. The present  
CC sequence represents cDNA encoding a human ovarian antigen of the  
CC invention. Note: The sequence data for this patent did not form part of  
CC the printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
XX SQ Sequence 632 BP; 145 A; 176 C; 201 G; 93 T; 0 U; 17 Other;  
SQ

Alignment Scores:  
Pred. No.: 3,186-40 Length: 632  
Score: 49.00 Matches: 49

Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 27.37% Indels: 0  
 DB: 6 Gaps: 0

US-09-701-674A-23 (1-179) x ABQ55427 (1-632)

QY 24 LysMetAspSerAsnLysGluPheValThrSerAsnGluSerThrGluGluSerSerAsn 43  
 DB 82 AAATGGATTCGAACAAGAAATTTGTACTTCCACGAGAGCACCGAGAGAGCTCCAAC 141  
 QY 44 CysGluAsnGlySerProGlnLysGlyArgGlyGlyLeuGlyLysArgGlyAlaPro 63  
 DB 142 TCGAGATGGGTCTCTCCAGAGAGGCGCGGCGCTTGGCAAGAGGAGGAGCGGCC 201  
 QY 64 ThrLysLysSerProLeuSerGlyVal 72  
 DB 202 ACCAAGAGAGCCCTCGAGCGGTCTC 228

RESULT 7  
 ACA56430  
 ID ACA56430 standard; cDNA; 240 BP.

XX AC ACA56430;

XX DT 06-JUN-2003 (first entry)

XX DE Chicken signalling pathway polynucleotide probe SEQ ID NO 1028.

XX KW Chicken; probe; ss; array element; Parkinson's disease;

XX KW signalling pathway population; cancer; adenocarcinoma; leukaemia;

XX KW immunopathy; AIDS; asthma; neuropathy; Alzheimer's disease; microarray.

XX OS Gallus gallus.

XX PN US6500938-B1.

XX PD 31-DEC-2002.

XX PF 30-JAN-1998; 98US-00016434.

XX PR 30-JAN-1998; 98US-00016434.

XX PA (INCY-) INCYTE GENOMICS INC.

XX PI Au-Young J, Seilhamer JJ;

XX DR WPI; 2003-352189/33.

XX PT Combination of polynucleotide probes, useful as array elements in a

XX PT microarray for monitoring the expression of a number of target

XX PT polynucleotides.

XX PS Claim 1; SEQ ID NO 1028; 65pp; English.

XX CC The invention relates to a combination which, comprises a number of  
 CC polynucleotide probes comprising a sequence selected from one of the 1490  
 CC sequences mentioned in the specification. The combination is useful as an  
 CC array element in a microarray for monitoring the expression of a number  
 CC of target polynucleotides. The microarray is particularly useful in the  
 CC diagnosis and treatment of cancer and immunopathology and neuropathology.  
 CC The microarray is useful in diagnostics and treatment regimens, drug  
 CC discovery and development, toxicological and carcinogenicity studies,  
 CC forensic and pharmacogenomics. The microarray is also useful for  
 CC monitoring progression of diseases and for developing sophisticated  
 CC profiles for the effects of currently available therapeutic drugs. The  
 CC combination is also useful for purifying a subpopulation of mRNAs, cDNAs  
 CC and genomic fragments and in research and diagnostic applications. The  
 CC array can detect changes in expression in a large number of genes coding  
 CC for different signalling pathway populations which can be used to diagnose  
 CC various diseases including cancer e.g. adenocarcinoma and leukaemia,  
 CC immunopathies e.g. AIDS and asthma, neuropathies e.g. Alzheimer's disease  
 CC and Parkinson's disease. The present sequence represents a polynucleotide

CC probe of the invention. Note: The sequence data for this patent did not  
 CC form part of the printed specification but was obtained in electronic  
 CC format directly from USPIO at  
 CC seqdata.uspto.gov/sequence.html?DocID=06500938B1

XX SQ Sequence 240 BP; 55 A; 83 C; 72 G; 27 T; 0 U; 3 Other;

Alignment Scores:  
 Pred. No.: 2,24e-27 Length: 240  
 Score: 36.00 Matches: 36  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 20.11% Indels: 0  
 DB: 10 Gaps: 0

US-09-701-674A-23 (1-179) x ACA56430 (1-240)

QY 90 AlaArgMetArgValLeuSerLysAlaPheSerArgLeuLysThrThrLeuProTrpVal 109  
 DB 90 GCCCGCATGCGAGTGTGAGCAGAGGCTTCTCCAGACTCAAGACCCCTGCGCTGGGTG 149

QY 110 ProProAspThrLysLysLeuSerLysLysLeuAspThrLeuArgLeuAlaSer 125

DB 150 CCCCCGACACCAAGCTCTCCAGCTGGACACGCTCAGGCTGGCGTCC 197

RESULT 8

AD156226  
 ID AD156226 standard; DNA; 240 BP.

XX AC AD156226;

XX DT 22-APR-2004 (first entry)

XX DE Human polynucleotide probe #1028.

XX KW Human; probe; ss; receptor-like polypeptide; transducing polypeptide;

XX KW effector-like polypeptide; cancer; immunopathology; neuropathology;

XX KW drug development; toxicology; carcinogenicity; bladder; bone;

XX KW signalling pathway polypeptide; adrenal gland; AIDS;

XX KW bone marrow; brain; breast; cervix; tumour; immunopathology; AIDS;

XX KW diabetes; pancreatitis; osteoporosis; ulcerative colitis; neuropathology;

XX KW dementia; amnesia; epilepsy; Alzheimer's disease; depression.

XX OS Homo sapiens.

XX PN US2004010136-A1.

XX PD 15-JAN-2004.

XX PF 26-NOV-2002; 2002US-00305720.

XX PR 30-JAN-1998; 98US-00016434.

XX PA (INCY-) INCYTE GENOMICS INC.

XX PI Au-Young J, Seilhamer JJ;

XX DR WPI; 2004-090520/09.

XX PT New composition comprising polynucleotide probes, useful as array

XX PT elements in a microarray for monitoring the expression of target

XX PT polynucleotides or purifying a subpopulation of mRNAs, cDNA, or genomic

XX PT fragments.

XX PS Claim 6; SEQ ID NO 1028; 73pp; English.

XX CC The invention relates to a composition of polynucleotide probes  
 CC comprising first polynucleotide probes comprising at least a portion of a  
 CC gene encoding a receptor-like polypeptide, second polynucleotide probes  
 CC comprising at least a portion of a gene encoding a transducing  
 CC polypeptide and third polynucleotide probes comprising at least a portion  
 CC of a gene encoding an effector-like polypeptide. The probes of the  
 CC composition are useful as array elements in a microarray for monitoring

the expression of target polynucleotides. The microarray is useful in the diagnosis and treatment of cancer, an immunopathology or a neuropathology. It can also be used for drug discovery and development, toxicological and carcinogenicity studies, forensics or pharmacogenomics. Microarrays can also be used for monitoring the progression of diseases that may be associated with the altered expression of signalling pathway polypeptides. The composition can also be used to purify a subpopulation of mRNAs, cDNAs, or genomic fragments in a sample. The expression profile is also useful for the diagnosis and treatment of cancer, e.g. cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast or cervix, an immunopathology, e.g. AIDS, diabetes, pancreatitis, osteoporosis or ulcerative colitis, or a neuropathology, e.g. dementia, amnesia, epilepsy, Alzheimer's disease or depression. This sequence represents a human polynucleotide probe of the invention. Note: The sequence data for this patent did not form part of the printed specification but was obtained in electronic format directly from USPTO at seqdata.uspto.gov/sequence.html.

Sequence 240 BP; 55 A; 83 C; 72 G; 27 T; 0 U; 3 Other;

Alignment Scores:  
Pred. No.: 2,24e-27 Length: 240  
Score: 36.00 Matches: 36  
Percent Similarity: 100.00% Conservativity: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 20.11% Indels: 0  
DB: 12 Gaps: 0

US-09-701-674A-23 (1-179) x ADI56226 (1-240)

Qy 90 AlaArgMetArgValLeuSerLysAlaPheSerArgLeuLysThrLeuProTyrVal 109  
Db 90 GCCCGCATCGAGTGTGAGCAAGCCCTCTCCAGACTCAAGACCACCTGCGGTG 149  
Qy 110 ProProAspThrLysLeuSerLysLeuAspThrLeuArgLeuAlaSer 125  
Db 150 CCCCCGACACCAAGCTCTCCAGCTGGACACGCTCAGCTGGCTGC 197

RESULT 9  
ACH23969  
ID ACH23969 standard; cDNA; 493 BP.  
AC ACH23969;  
DT 13-OCT-2003 (first entry)  
DE Human adult ovary cDNA #2349.  
XX Human; ss; sequencing by hybridisation; SBH; expressed sequence tag; EST;  
KW genome mapping; biodiversity; genetic disorder.  
XX Homo sapiens.

US2003073623-A1.  
17-APR-2003.  
30-JUL-2001; 2001US-00918995.  
30-JUL-2001; 2001US-00918995.  
(DRMA/) DRMANAC R T.  
(LABA/) LABAT I.  
(STAC/) STACHE-CRAIN B.  
(DICK/) DICKSON M C.  
(JONE/) JONES L W.

Drmanac RT, Labat I, Stache-Crain B, Dickson MC, Jones LW;  
WPI; 2003-615964/58.  
New polynucleotide sequences obtained from various cDNA libraries, useful as hybridization probes, as oligomers for PCR, for chromosome and gene

mapping, in the recombinant production of protein, or in generating antisense DNA or RNA.

Claim 1; SEQ ID NO 11181; 44pp; English.

The invention relates to an isolated polynucleotide comprising any one of 38043 cDNA sequences, appearing as ACH12789-ACH50831, whose sequence was determined by the technique of SBH (sequencing by hybridisation). Also included is a purified polypeptide comprising a sequence corresponding to a reading frame of the novel polynucleotide. The nucleic acid sequences are useful in diagnostics as expressed sequence tags (EST) for identifying expressed genes or for physical mapping of the human genome, in forensics, in assessing biodiversity, or in identifying mutations responsible for genetic disorders and other traits. The nucleotide sequences are also useful as hybridisation probes, as oligomers for PCR, for chromosome and gene mapping, in the recombinant production of protein, or in generating antisense DNA or RNA. The purified polypeptide is useful for generating antibodies specific for it. The present sequence is one of the 38043 isolated cDNA/EST sequences. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from USPTO at seqdata.uspto.gov/sequence.html?DocID=20030073623

Sequence 493 BP; 121 A; 136 C; 121 G; 112 T; 0 U; 3 Other;

Alignment Scores:  
Pred. No.: 6.28e-19 Length: 493  
Score: 28.00 Matches: 28  
Percent Similarity: 100.00% Conservativity: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 15.64% Indels: 0  
DB: 9 Gaps: 0

US-09-701-674A-23 (1-179) x ACH23969 (1-493)

Qy 152 TrpProPheMetValAlaGlyLysProGluSerAspLeuLysGluValValThrAlaSer 171  
Db 102 TGGCCCTTTAATGTTGGCCGGAAACCCGAGAGTACCTGAAGAAGTGTGACCGGAGC 161  
Qy 172 ArgLeuCysGlyThrThrAlaSer 179  
Db 162 CGCTTATGTGAACACACCGCGTCC 185

RESULT 10  
AAZ50465  
ID AAZ50465 standard; cDNA; 1716 BP.  
AC AAZ50465;  
DT 23-MAY-2000 (first entry)  
DE Human muscudin bHLH protein encoding cDNA.

Muscudin; basic helix-loop-helix; bHLH; transcription factor; myogenesis; screening; myogenic disease; aberration; muscle development; cytostatic; gene therapy; diagnostic agent; muscular dystrophy; myopathy; neuromuscular; skeletomuscular; myogenic cancer; human; ss.

Homo sapiens.

Key Location/Qualifiers  
CDS 2..607  
/tag= a  
/product= "Human muscudin bHLH protein"  
/note= "Lacks Alanine-Threonine myogenic recognition motif"

WO200006720-A1.

10-FEB-2000.

30-JUL-1999; 99WO-AU000623.



```

XX WPI; 2003-875398/81.
DR P-PSDB; ADE25769.
XX
XX Combination containing several polynucleotide that are differentially
PT expressed in foam cells and complements of the polynucleotides, useful
PT for diagnosing cardiovascular disease or atherosclerosis.
XX
XX Claim 1; SEQ ID NO 97; 37pp; English.
XX
XX The invention relates to a combination comprising several polynucleotides
CC having any one of 127 sequences (S1) such as the sequence of human
CC calmodulin gene, human mRNA for KIAA0930 protein, leukotriene A4
CC hydrolase, human CGI-142 protein mRNA, human K+ channel beat 2 subunit
CC mRNA, etc., and their complements. The cDNAs are differentially expressed
CC in LPS (lipopolysaccharide)-treated foam cells. Also included are
CC obtaining an extended or full length gene from a library of nucleic acid
CC sequences, an expression vector containing the nucleic acids, a host cell
CC containing the vector, a purified polypeptide appearing as ADE25750 and
CC ADE25751, producing a protein by culturing the host cell, and a
CC composition comprising a purified antibody that specifically binds to the
CC proteins. The foam cell-expressed nucleic acids are useful for a high
CC throughput detection of differential expression of one or more
CC polynucleotides in a sample. The sample is from a subject with
CC atherosclerosis and comparison with a standard defines early, mid or late
CC stages of the disorder. The foam cell-expressed nucleic acids are useful
CC for high throughput screening of a library of molecules or compounds to
CC identify a ligand which binds a polynucleotide. The library is chosen
CC from DNA molecules, peptides, proteins and RNA molecules. The protein is
CC useful for a high throughput screening of library of molecules or
CC compounds to identify at least one ligand which specifically binds a
CC protein, for purifying a ligand from a sample for making an antibody. The
CC foam cell-expressed nucleic acids are useful for diagnosing
CC cardiovascular disorder. The foam cell-expressed nucleic acids are useful
CC as elements on a microarray which can be used for detecting related
CC polynucleotide in a sample, diagnosing cardiovascular disease,
CC atherosclerosis. The present sequence represents a cDNA whose expression
CC is upregulated in LPS treated foam cells.
XX
SQ Sequence 2177 BP; 423 A; 691 C; 650 G; 413 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 2,576-18 Length: 2177
Score: 28.00 Matches: 28
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 15.64% Indels: 0
DB: 10 Gaps: 0

US-09-701-674A-23 (1-179) x ADE25693 (1-2177)
QY 106 LeuProTTPValProProAspThrLysLeuSerLysLeuAspThrLeuArgLeuAlaSer 125
DB 857 CTGCCCTGGTGGTCCCGGACACTAAGCTCTCCAGCTGACACGCTCGGCTGCTTCC 916

QY 126 SerTyrlleAlaHisLeuArgGln 133
DB 917 AGTTACATCGCTCACCTGGCGCAG 940

RESULT 13
AAA72428
ID AAA72428 standard; cDNA; 2196 BP.
XX
XX AAA72428;
AC
AC
DI 19-DEC-2000 (first entry)
XX
DE Human nucleic acid-binding protein NuABP-47 cDNA.
XX
XX Human nucleic acid-binding protein; NuABP; agonist; antagonist; EST;
KW expressed sequence tag; drug screening; recombinant expression; antibody;
KW reproductive disorder; infertility; immunological disorder;
KW neurological disorder; cell proliferative disorder; cancer; tumour; ss.

```

```

XX Homo sapiens.
OS
PN WC2000044900-A2.
XX
XX 03-AUG-2000.
PD
XX
XX 28-JAN-2000; 2000MO-US002237.
PF
XX
XX 29-JAN-1999; 99US-0117904P.
PR
XX
XX 29-JAN-1999; 99US-0117905P.
PR
XX
XX (INCY-) INCYTE PHARM INC.
PA
XX
XX Tang YT, Lal P, Hillman JL, Yue H, Azimzai Y, Lu AMD, Baughn MR;
PI Tran B, Shih LL, Au-Young JL;
XX
XX WPI: 2000-499332/44.
DR
XX
XX P-PSDB; AAB21043.
DR
XX
XX Novel nucleic acid binding proteins, used to identify agonists and
PT antagonists of them, for the treatment of reproductive, immunological,
PT neurological and cell proliferative disorders including cancer.
XX
XX Claim 4; Page 176; 180pp; English.
XX
XX Sequences AAA72382-A72436 represent cDNAs encoding novel human nucleic
CC acid-binding proteins (NuABPs; AAB20997-B21051). These cDNAs were
CC produced by extension from an appropriate EST (expressed sequence tag)
CC using primers designed using the EST. The invention also relates to
CC expression constructs, host cells and transgenic organisms comprising a
CC human NuABP nucleic acid, recombinant production of the human NuABPs, and
CC antibodies against the human NuABPs, and also to methods of screening
CC modulators of human NuABP activity or expression. The human NuABPs, and
CC their agonists and antagonists are used to treat diseases associated with
CC overexpression or underexpression of functional NuABPs. Human NuABP
CC proteins and nucleotides, and NuABP agonists and antagonists can be used
CC to diagnose, treat and prevent reproductive, immunological, neurological
CC and cell proliferative disorders. Reproductive disorders that may be
CC treated using compositions of the invention include infertility,
CC endometriosis, disruptions of the menstrual cycle and disruptions of
CC spermatogenesis. Immunological disorders that may be treated include
CC AIDS, allergies, and autoimmune disorders such as multiple sclerosis,
CC rheumatoid arthritis, diabetes and systemic lupus erythematosus.
CC Neurological disorders that may be treated include epilepsy,
CC neurodegenerative conditions such as Alzheimer's disease and Parkinson's
CC disease, prion diseases such as Creutzfeldt-Jakob disease, and mental
CC disorders such as schizophrenia. Cell proliferative disorders that may be
CC treated include a wide variety of cancers, and also arteriosclerosis,
CC atherosclerosis, cirrhosis and psoriasis
XX
SQ Sequence 2196 BP; 438 A; 691 C; 653 G; 414 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 2,596-18 Length: 2196
Score: 28.00 Matches: 28
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 15.64% Indels: 0
DB: 3 Gaps: 0

US-09-701-674A-23 (1-179) x AAA72428 (1-2196)
QY 106 LeuProTTPValProProAspThrLysLeuSerLysLeuAspThrLeuArgLeuAlaSer 125
DB 857 CTGCCCTGGTGGTCCCGGACACTAAGCTCTCCAGCTGACACGCTCGGCTGCTTCC 916

QY 126 SerTyrlleAlaHisLeuArgGln 133
DB 917 AGTTACATCGCTCACCTGGCGCAG 940

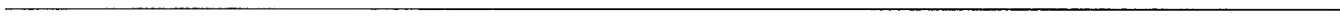
RESULT 14
ADJ56314

```





Search completed: December 17, 2004, 01:22:05  
Job time : 461 secs



**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 16, 2004, 19:03:27 ; Search time 728 Seconds

(without alignments)  
9646.299 Million cell updates/sec

Title: US-09-701-674A-54

Perfect score: 1272

Sequence: 1 tctacggccagcactctggg.....attgctataaaaaaaaaa 1272

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 4093002 seqs, 2760418825 residues

Total number of hits satisfying chosen parameters: 8165004

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications\_NA.\*

- 1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq.\*
- 2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq.\*
- 3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq.\*
- 4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq.\*
- 5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq.\*
- 6: /cgn2\_6/ptodata/1/pubpna/PCTUS\_PUBCOMB.seq.\*
- 7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq.\*
- 8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq.\*
- 9: /cgn2\_6/ptodata/1/pubpna/US09A\_PUBCOMB.seq.\*
- 10: /cgn2\_6/ptodata/1/pubpna/US09B\_PUBCOMB.seq.\*
- 11: /cgn2\_6/ptodata/1/pubpna/US09C\_PUBCOMB.seq.\*
- 12: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq.\*
- 13: /cgn2\_6/ptodata/1/pubpna/US10A\_PUBCOMB.seq.\*
- 14: /cgn2\_6/ptodata/1/pubpna/US10B\_PUBCOMB.seq.\*
- 15: /cgn2\_6/ptodata/1/pubpna/US10C\_PUBCOMB.seq.\*
- 16: /cgn2\_6/ptodata/1/pubpna/US10D\_PUBCOMB.seq.\*
- 17: /cgn2\_6/ptodata/1/pubpna/US10E\_PUBCOMB.seq.\*
- 18: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq.\*
- 19: /cgn2\_6/ptodata/1/pubpna/US11\_NEW\_PUB.seq.\*
- 20: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq.\*
- 21: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	581	45.7	697	9	US-09-954-456-829
2	581	45.7	697	9	Sequence 829, App
3	581	45.7	697	10	Sequence 1245, App
4	581	45.7	697	10	Sequence 1028, App
5	525.6	41.3	632	16	Sequence 679, App
6	466.2	36.7	471	9	Sequence 1307, App
7	466.2	36.7	471	9	Sequence 881, App
8	466.2	36.7	471	9	Sequence 1431, App
9	466.2	36.7	471	10	Sequence 17, App
10	434.2	34.1	493	10	Sequence 10, App
11	297.2	23.4	482	10	Sequence 1181, A
12	213	16.7	240	16	Sequence 9069, App
					Sequence 1028, App

13	211.6	16.6	2177	15	US-10-247-671-97	Sequence 97, Appl
14	211.6	16.6	2196	15	US-10-084-817-120	Sequence 120, App
15	210	16.5	1716	17	US-10-775-169-201	Sequence 201, App
16	208.6	16.4	2382	15	US-10-240-965-83	Sequence 83, Appl
17	198	15.6	1939	17	US-10-755-889-365	Sequence 365, App
18	81.8	6.4	446	9	US-09-833-381-505	Sequence 505, App
19	78.4	6.2	1140	17	US-10-450-826-124	Sequence 124, App
20	71.8	5.6	609	15	US-10-029-386-22740	Sequence 22740, A
21	71.8	5.6	1800	10	US-09-771-357-106	Sequence 106, App
22	71.8	5.6	1800	15	US-10-059-579-106	Sequence 106, App
23	71	5.6	450	13	US-10-004-717-22	Sequence 22, Appl
24	71	5.6	450	18	US-10-860-373-22	Sequence 22, Appl
25	71	5.6	450	18	US-10-860-724-22	Sequence 22, Appl
26	69.4	5.5	1457	9	US-09-954-531-982	Sequence 982, App
27	69.4	5.5	1467	16	US-10-062-674-1950	Sequence 1950, App
28	69.4	5.5	1678	15	US-10-240-965-119	Sequence 119, App
29	69	5.4	158405	14	US-10-175-523-86	Sequence 86, Appl
30	68.8	5.4	2537	16	US-10-466-164-25	Sequence 25, Appl
31	68.8	5.4	2550	18	US-10-335-053-106	Sequence 106, App
32	68.8	5.4	2601	9	US-09-880-107-2433	Sequence 2433, App
33	68.2	5.4	82660	13	US-10-087-192-2017	Sequence 2017, App
34	68.2	5.4	83493	18	US-10-331-059-61	Sequence 61, Appl
35	68	5.3	651	9	US-09-749-728B-22	Sequence 22, Appl
36	67.6	5.3	1132	9	US-09-778-844-23	Sequence 23, Appl
37	67.6	5.3	256525	13	US-10-087-192-451	Sequence 451, App
38	67.4	5.3	444	18	US-10-674-124A-6324	Sequence 6324, App
39	67.4	5.3	2466	13	US-10-044-090-251	Sequence 251, App
40	66.8	5.3	486	10	US-09-918-995-22633	Sequence 22633, A
41	66.2	5.2	505	10	US-09-918-995-25117	Sequence 25117, A
42	66.2	5.2	42999	9	US-09-799-462A-17	Sequence 17, Appl
43	66.2	5.2	42999	10	US-09-836-911A-17	Sequence 17, Appl
44	66.2	5.2	42999	10	US-09-738-630-73	Sequence 73, Appl
45	66.2	5.2	42999	13	US-10-125-767-17	Sequence 17, Appl

ALIGNMENTS

RESULT 1

US-09-954-456-829  
; Sequence 829, Application US/09954456  
; Patent No. US20020115057A1  
; GENERAL INFORMATION:  
; APPLICANT: Young, Paul  
; TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using Canc  
; TITLE OF INVENTION: Sets  
; FILE REFERENCE: 689290-76  
; CURRENT APPLICATION NUMBER: US/09/954, 456  
; CURRENT FILING DATE: 2001-09-18  
; PRIOR APPLICATION NUMBER: US/60/233, 617  
; PRIOR FILING DATE: 2000-09-18  
; PRIOR APPLICATION NUMBER: US/60/234, 052  
; PRIOR FILING DATE: 2000-09-20  
; PRIOR APPLICATION NUMBER: US/60/234, 923  
; PRIOR FILING DATE: 2000-09-25  
; PRIOR APPLICATION NUMBER: US/60/235, 134  
; PRIOR FILING DATE: 2000-09-25  
; PRIOR APPLICATION NUMBER: US/60/235, 637  
; PRIOR FILING DATE: 2000-09-26  
; PRIOR APPLICATION NUMBER: US/60/235, 638  
; PRIOR FILING DATE: 2000-09-26  
; PRIOR APPLICATION NUMBER: US/60/235, 711  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: US/60/235, 720  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: US/60/235, 840  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: US/60/235, 863  
; PRIOR FILING DATE: 2000-09-27  
; NUMBER OF SEQ ID NOS: 2276  
; SOFTWARE: Patentin version 3.0  
; SEQ ID NO 829  
; LENGTH: 697

; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; OTHER INFORMATION: n=a,t,g or c  
US-09-954-456-829

Query Match 45.7%; Score 581; DB 9; Length 697;  
Best Local Similarity 95.1%; Pred. No. 4.4e-132;  
Matches 621; Conservative 0; Mismatches 27; Indels 5; Gaps 2;

```
QY 620 TGGACACGCTCAGGCTGGGCTCAGCTACATCGCCCACTTGGGAGAGATCTGGCTAACG 679
DB 1 TGGACACGCTCAGGCTGGGCTCAGCTACATCGCCCACTTGGGAGAGATCTGGCTAACG 60
QY 680 ACAATACAGAACGGGTACATTCACCGGTCAACCTGACGTGGCCCTTTATGGTGGCG 739
DB 61 ACAATACAGAACGGGTACATTCACCGGTCAACCTGACGTGGCCCTTTATGGTGGCG 120
QY 740 GGAACCCGAGAGTGACCTGAAGAGTGGTGACCGCGAGCGCTTATGTGAAACCCG 799
DB 121 GGAACCCGAGAGTGACCTGAAGAGTGGTGACCGCGAGCGCTTATGTGAAACCCG 180
QY 800 CGTCTGACCTTGGAGTGGAGTCTGGGAAAGGCGCGCTCCCGGGGGA-GCGGGCCCC 858
DB 181 CGTCTGACCTTGGAGTGGAGTCTGGGAAAGGCGCGCTCCCGGGGANGCGGNCNT 240
QY 859 GGAAGGCGACCCCTGCCCTCAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 918
DB 241 GGAAGGCGACCCCTGCCCTCAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 300
QY 919 TCTCTGTCCACCCCGCGAGAACACTTTACACGACGAGGAGATTCGTTTCCAAACCCAGA 978
DB 301 TCTCTGTCCACCCCGCGAGAACACTTTACACGACGAGGAGATTCGTTTCCAAACCCAGA 360
QY 979 GGAGATCAATTTGATCAAAAGATCCCATCTATTAACTTTTAACTTTTAACTTTTAACTTT 1038
DB 361 GGAGATCAATTTGATCAAAAGATCCCATCTATTAACTTTTAACTTTTAACTTTTAACTTT 420
QY 1039 ATGACTCTGCAAGCCTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1098
DB 421 ATGACTCTGCAAGCCTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 480
QY 1099 AAGAGCTATCAATGTATCTTTTGTACAAATATGTTGTAAAGTGTAGATCATAGGATGCT 1158
DB 481 AAGAGCTATCAATGTATCTTTTGTACAAATATGTTGTAAAGTGTAGATCATAGGATGCT 540
QY 1159 GACTTTGACAGTCACATTTATAAAGTAACTTCACTTAAAGATATATTTTTTCAACAA 1218
DB 541 GACTTTGACAGTCACATTTATAAAGTAACTTCACTTAAAGATATATTTTTTCAACAA --- 597
QY 1219 GTTTTGTCTACTTTTGAATAAATCTTTCTTTATATGCTTAAATAAAAAAAAA 1271
DB 598 -AGTTTGCACCTTTTGAATAAATCTTTCTTTATATGCTTAAATAAAAAAAAAAGA 649
```

## RESULT 2

US-09-954-456-1245  
; Sequence 1245, Application US/09954456  
; Patent No. US20020115057A1  
; GENERAL INFORMATION:  
; APPLICANT: Young, Paul  
; TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using Canc  
; FILE OF INVENTION: Sets  
; FILE REFERENCE: 689290-76  
; CURRENT APPLICATION NUMBER: US/09/954,456  
; CURRENT FILING DATE: 2001-09-18  
; PRIOR APPLICATION NUMBER: US/60/233,617  
; PRIOR FILING DATE: 2000-09-18  
; PRIOR APPLICATION NUMBER: US/60/234,052  
; PRIOR FILING DATE: 2000-09-20  
; PRIOR APPLICATION NUMBER: US/60/234,923  
; PRIOR FILING DATE: 2000-09-25

; PRIOR APPLICATION NUMBER: US/60/235,134  
; PRIOR FILING DATE: 2000-09-25  
; PRIOR APPLICATION NUMBER: US/60/235,637  
; PRIOR FILING DATE: 2000-09-26  
; PRIOR APPLICATION NUMBER: US/60/235,638  
; PRIOR FILING DATE: 2000-09-26  
; PRIOR APPLICATION NUMBER: US/60/235,711  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: US/60/235,720  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: US/60/235,840  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: US/60/235,863  
; PRIOR FILING DATE: 2000-09-27  
; NUMBER OF SEQ ID NOS: 2276  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 1245  
; LENGTH: 697  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; OTHER INFORMATION: n=a,t,g or c  
US-09-954-456-1245

Query Match 45.7%; Score 581; DB 9; Length 697;  
Best Local Similarity 95.1%; Pred. No. 4.4e-132;  
Matches 621; Conservative 0; Mismatches 27; Indels 5; Gaps 2;

```
QY 620 TGGACACGCTCAGGCTGGGCTCAGCTACATCGCCCACTTGGGAGAGATCTGGCTAACG 679
DB 1 TGGACACGCTCAGGCTGGGCTCAGCTACATCGCCCACTTGGGAGAGATCTGGCTAACG 60
QY 680 ACAATACAGAACGGGTACATTCACCGGTCAACCTGACGTGGCCCTTTATGGTGGCG 739
DB 61 ACAATACAGAACGGGTACATTCACCGGTCAACCTGACGTGGCCCTTTATGGTGGCG 120
QY 740 GGAACCCGAGAGTGACCTGAAGAGTGGTGACCGCGAGCGCTTATGTGAAACCCG 799
DB 121 GGAACCCGAGAGTGACCTGAAGAGTGGTGACCGCGAGCGCTTATGTGAAACCCG 180
QY 800 CGTCTGACCTTGGAGTGGAGTCTGGGAAAGGCGCGCTCCCGGGGGA-GCGGGCCCC 858
DB 181 CGTCTGACCTTGGAGTGGAGTCTGGGAAAGGCGCGCTCCCGGGGANGCGGNCNT 240
QY 859 GGAAGGCGACCCCTGCCCTCAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 918
DB 241 GGAAGGCGACCCCTGCCCTCAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 300
QY 919 TCTCTGTCCACCCCGCGAGAACACTTTACACGACGAGGAGATTCGTTTCCAAACCCAGA 978
DB 301 TCTCTGTCCACCCCGCGAGAACACTTTACACGACGAGGAGATTCGTTTCCAAACCCAGA 360
QY 979 GGAGATCAATTTGATCAAAAGATCCCATCTATTAACTTTTAACTTTTAACTTTTAACTTT 1038
DB 361 GGAGATCAATTTGATCAAAAGATCCCATCTATTAACTTTTAACTTTTAACTTTTAACTTT 420
QY 1039 ATGACTCTGCAAGCCTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1098
DB 421 ATGACTCTGCAAGCCTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 480
QY 1099 AAGAGCTATCAATGTATCTTTTGTACAAATATGTTGTAAAGTGTAGATCATAGGATGCT 1158
DB 481 AAGAGCTATCAATGTATCTTTTGTACAAATATGTTGTAAAGTGTAGATCATAGGATGCT 540
QY 1159 GACTTTGACAGTCACATTTATAAAGTAACTTCACTTAAAGATATATTTTTTCAACAA 1218
DB 541 GACTTTGACAGTCACATTTATAAAGTAACTTCACTTAAAGATATATTTTTTCAACAA --- 597
QY 1219 GTTTTGTCTACTTTTGAATAAATCTTTCTTTATATGCTTAAATAAAAAAAAA 1271
DB 598 -AGTTTGCACCTTTTGAATAAATCTTTCTTTATATGCTTAAATAAAAAAAAAAGA 649
```

RESULT 3  
US-09-760-706-1028  
; Sequence 1028, Application US/09960706  
; Publication No. US20030134280A1  
; GENERAL INFORMATION:  
; APPLICANT: Munger, William E.  
; TITLE OF INVENTION: Identifying Drugs for and Diagnosis of Benign Prostatic Hyperplasia  
; TITLE OF INVENTION: Gene Expression Profiles  
; FILE REFERENCE: 44921-5029-01US  
; CURRENT APPLICATION NUMBER: US/09/960,706  
; CURRENT FILING DATE: 2001-09-24  
; PRIOR APPLICATION NUMBER: 60/223,323  
; PRIOR FILING DATE: 2000-08-07  
; PRIOR APPLICATION NUMBER: 05/873,319  
; PRIOR FILING DATE: 2001-06-05  
; NUMBER OF SEQ ID NOS: 1124  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1028  
; LENGTH: 697  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (1)..(697)  
; OTHER INFORMATION: n = a or c or g or t  
; US-09-760-706-1028

Query Match 45.7%; Score 581; DB 10; Length 697;  
Best Local Similarity 95.1%; Pred. No. 4.4e-132;  
Matches 621; Conservative 0; Mismatches 27; Indels 5; Gaps 2;

Qy	620	TGACACGCTCAGCTGGCGTCCAGCTACATCCCCACTTGAGCGCAGATCCTGGCTAACG	679
Db	1	TGACACGCTCAGCTGGCGTCCAGCTACATCCCCACTTGAGCGCAGATCCTGGCTAACG	60
Qy	680	ACAAATACGAGAACGGGTACATTCACCGGTCAACTGACGTGGCGCTTTATGGTGGCGG	739
Db	61	ACAAATACGAGAACGGGTACATTCACCGGTCAACTGACGTGGCGCTTTATGGTGGCGG	120
Qy	740	GGAAACCCGAGAGTGACCTGAAGAGTGTGACCGGAGCCGCTTATGTGGAACACCG	799
Db	121	GGAAACCCGAGAGTGACCTGAAGAGTGTGACCGGAGCCGCTTATGTGGAACACCG	180
Qy	800	CGTCCTGACCTTGGAGTGGCGAGTCTGGGAAAGCGCGCTCCCGGGGGGA-GCGGGGCC	858
Db	181	CGTCCTGACCTTGGAGTGGCGAGTCTGGGAAAGCGCGCTCCCGGGGGGANGCNCNC	240
Qy	859	GGGAAGCGACCCCTGCCTCAGTCTCTGTCTGTCTGTCTTCCCTCGCAATGCTCCTC	918
Db	241	GGGAAGCGACCCCTGCCTCAGTCTCTGTCTGTCTGTCTTCCCTCGCAATGCTCCTC	300
Qy	919	TCTCTGTCCACCCCGGAGAACACTTTACACGACGAGGAGATTGTTTCCAAACCAGA	978
Db	301	TCTCTGTCCACCCCGGAGAACACTTTACACGACGAGGAGATTGTTTCCAAACCAGA	360
Qy	979	GGAGATCAATTGTACTTACAAAGATCCCAATCTATTAACTTTTAACTTACCGTGA	1038
Db	361	GGAGATCAATTGTACTTACAAAGATCCCAATCTATTAACTTTTAACTTACCGTGA	420
Qy	1039	ATGACTCTGCAAGCCTTGCTGGTCCAAAGTCCCAATATGTAAATTAATAATATATAAT	1098
Db	421	ATGACTCTGCAAGCCTTGCTGGTCCAAAGTCCCAATATGTAAATTAATAATATATAAT	480
Qy	1099	AAGAGCCTATCAATGTATCTTTGTACAAATATGTTGTAAGTGTAGATCATAGATAGCT	1158
Db	481	AAGAGCCTATCAATGTATCTTTGTACAAATATGTTGTAAGTGTAGATCATAGATAGCT	540
Qy	1159	GACTTTGACAGTCACATTTATAAGTAATTCACCTTAAAGATATATATTTTTTCAAACA	1218
Db	541	GACTTTGACAGTCACATTTATAAGTAATTCACCTTAAAGATATATATTTTTTCAAACA	597

```

Qy 1219 GTTTTGGCTACTTTGAAATAAATCTTTCTTTATATGTCTAAAAAAGAAA 1271
Db 598 -AGTTTGCACTTTTGAATAAACCTTCTTTATATGTCTAAAAAAGAAA 649

RESULT 4
US-09-873-319-679
; Sequence 679, Application US/09873319A
; Publication No. US20030134324A1
; GENERAL INFORMATION:
; APPLICANT: Munger, William E.
; APPLICANT: Kulkarni, Prakash
; APPLICANT: Getzenberg, Robert H.
; APPLICANT: Waga, Iwao
; APPLICANT: Yamamoto, Jun
; TITLE OF INVENTION: Identifying Drugs for and Diagnosis of Benign Prostatic
; TITLE OF INVENTION: Hyperplasia Using Gene Expression Profiles
; FILE REFERENCE: 44921-5029-US
; CURRENT APPLICATION NUMBER: US/09/873,319A
; CURRENT FILING DATE: 2001-06-05
; EARLIER APPLICATION NUMBER: US 60/223,323
; EARLIER FILING DATE: 2000-08-07
; NUMBER OF SEQ ID NOS: 755
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 679
; LENGTH: 697
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20030134324A1 W73859
; FEATURE:
; LOCATION: (1)..(697)
; OTHER INFORMATION: n = a or c or g or t
US-09-873-319-679

Query Match 45.7%; Score 581; DB 10; Length 697;
Best Local Similarity 95.1%; Pred. No. 4.4e-132;
Matches 621; Conservative 0; Mismatches 27; Indels 5; Gaps 2;

Qy 620 TGGACAGCTCAGCGTGGCGTCAGCTACATCCGCCCATTTAGGCAGATCCTGGCTAACG 679
Db 1 TGGACAGCTCAGCGTGGCGTCAGCTACATCCGCCCATTTAGGCAGATCCTGGCTAACG 60

Qy 680 ACAAATACGAGAAGGGGTACAATTACCCGGTCAACCTGAGCTGGCCCTTTATGGTGGCG 739
Db 61 ACAAATACGAGAAGGGGTACAATTACCCGGTCAACCTGAGCTGGCCCTTTATGGTGGCG 120

Qy 740 GGAAACCCGAGAGTGACCTGAAAGAGTGTGAGACGGGAGCGCTTATGTGGAACCAACG 799
Db 121 GGAAACCCGAGAGTGACCTGAAAGAGTGTGAGACGGGAGCGCTTATGTGGAACCAACG 180

Qy 800 CGTCTCTGACCTTGGAGTGGAGTCTTGGGAAAGGCGGCTCCCGGGGGGA-GGGGGCCCC 858
Db 181 CGTCTCTGACCTTGGAGTGGAGTCTTGGGAAAGGCGGCTCCCGGGGGGANGCGCNCT 240

Qy 859 GGAAGCGCACCCCTGCCCTCAGTGCTCTCTGTCTCTGTCTTCCCCTTCGCAATGCTCCTC 918
Db 241 GGAAGCGCACCCCTGCCCTCAGTGCTCTCTGTCTCTGTCTTCCCCTTCGCAATGCTCCTC 300

Qy 919 TCTCTGTCCACCCCGGAGAACACTTTACAACGAGGAGGATTCGTTTTCCAAACCAGA 978
Db 301 TCTCTGTCCACCCCGGAGAACACTTTACAACGAGGAGGATTCGTTTTCCAAACCAGA 360

Qy 979 GGAGATCAATTGTACTTTACAAAGATTCCTTAATCTTTAACTTTTACCTTACCGTGA 1038
Db 361 GGAGATCAATTGTACTTTACAAAGATTCCTTAATCTTTAACTTTTACCTTACCGTGA 420

Qy 1039 ATGACTCTGAAGCCTTGCTGGTCCAAGTCSAATATGTAATATAAATATATAAATAGAT 1098
Db 421 ATGACTCTGAAGCCTTGCTGGTCCAAGTCSAATATGTAATATAAATATAAATATAGAT 480

Qy 1099 AAGAGCCTATCAATGTATCTTTTGTACAAATGTTGTAATAATGTGTAATAATAGATCATAGATAGCT 1158

```

```

Db 481 AGAGCGTATCAATGTAATCTTTTGTACAAATATGTTGTAATAATGATCATAGATAGCT 540
QY 1159 GACTTTGACAGTCACATTATTAAGTAATTCACITTAAGATATATATTTTCAACAA 1218
Db 541 GACTTTGACAGTCACATTATTAAGTAATTCACITTAAGATATATATTTTCAACAA 597
QY 1219 GTTTTGCTACTTTTGAATAAATCTTTCTTTATATATGCTTAAAAA 1271
Db 598 -AGTTTGCATCTTTTGAATAAATCTTTCTTTATATGCTTAAAAA 649

```

## RESULT 5

```

US-10-264-049-1307
; Sequence 1307, Application US/10264049
; Publication No. US20040005579A1
; GENERAL INFORMATION:
; APPLICANT: Birse et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PA133P1
; CURRENT APPLICATION NUMBER: US/10/264,049
; CURRENT FILING DATE: 2002-10-04
; PRIOR APPLICATION NUMBER: PCT/US01/18569
; PRIOR FILING DATE: 2001-06-07
; PRIOR APPLICATION NUMBER: US 60/209,467
; PRIOR FILING DATE: 2000-06-07
; NUMBER OF SEQ ID NOS: 4360
; SOFTWARE: PatentIn Ver. 3.1
; SEQ ID NO 1307
; LENGTH: 632

```

```

; TYPE: DNA
; ORGANISM: Homo sapiens
; NAME/KEY: misc feature
; LOCATION: (78)..(79)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (174)..(174)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (231)..(231)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (249)..(249)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (263)..(263)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (265)..(265)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (280)..(280)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (331)..(331)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (433)..(433)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (467)..(467)
; OTHER INFORMATION: n equals a,t,g, or c

```

```

; NAME/KEY: misc feature
; LOCATION: (498)..(498)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (537)..(537)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (546)..(546)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (552)..(552)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (623)..(623)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (629)..(629)
; OTHER INFORMATION: n equals a,t,g, or c
; US-10-264-049-1307

Query Match 41.3%; Score 525.6; DB 16; Length 632;
Best Local Similarity 93.4%; Pred. No. 1.6e-118;
Matches 593; Conservative 0; Mismatches 36; Indels 6; Gaps 5;

QY 257 CTCCTCCCAACATGTCACCGGCTCCTCAGCGATGTGGAGACCTTCAAGAGTGGAGA 316
Db 2 CTCCTCCCAACATGTCACCGGCTCCTCAGCGATGTGGAGACCTTCAAGAGTGGAGA 61
QY 317 TGTGGAATGTGACGGGTTGAAAATGGAATCGAAACAAGAAATTTGTGACTTCAACGAGA 376
Db 62 TGTGGAATGTGACGGGNTGAAAATGGAATCGAAACAAGAAATTTGTGACTTCAACGAGA 121
QY 377 GCACCGAGGAGCTCCAACTGGGAGATGGGTCTCCCGAGAGGGCGCGCGCCCTGG 436
Db 122 GCACCGAGGAGCTCCAACTGGGAGATGGGTCTCCCGAGAGGGCGCGCGCCCTGG 181
QY 437 GCAAGAGAGGAGGCGCGCCCAACAAGAGAGGCGGCTCAGCGGGTCAAGCAGGAGGGA 496
Db 182 GCAAGAGAGGAGGCGCGCCCAACAAGAGAGGCGGCTCAGCAGGAGGGA 241
QY 497 AGCAGTCCAGGCGCAACGCGCGCAGAGAGGGGCGCGCATGCGAGTCTGAGCA 556
Db 242 AGCAGTNCAGCGCAACGCGCGCAGAGAGGGGCGCGCATGCGAGTCTGAGCA 301
QY 557 AGGCTTTCTCCAGACTCAAGACACACCTCCTGGGTGCCCGCCCGACCAAGCTCTCA 616
Db 302 AGGCTTTCTCCAGACTCAAGACACACCTCCTGGGTGCCCGCCCGACCAAGCTCTCA 361
QY 617 AGCTGACACGCTCAGCTGGGCTCAGCTACATGCCCACTTGAAGCAGATCTTGCTA 676
Db 362 AGCTGACACGCTCAGCTGGGCTCAGCTACATGCCCACTTGAAGCAGATCTTGCTA 421
QY 677 AGCACAATACGAGACGCGGTACATTCACCGGTCAACCTGAGTGGCGCTTTATGGTG 736
Db 422 ACACCAATACNAGAACGCGGTACATTCACCGGG--CAACTGACGTGGNCTTTATGGTG 479
QY 737 CCGGGAACCCGAGAGTGACCTGAAAGAGAGTGG-TGACCGCAGCGCGCTTATGTGAACC 795
Db 480 -CGGGAACCCGAGAGTGACCTGAAAGAGAGTGGTTGACCGCAGCGCGGTATGTGAANC 538
QY 796 ACCGCTCTGACCTTGGAGTGGCGGTGCGAGTCTGGAAGAGCGGCTCCCGGGGAGCGGC 855
Db 539 ACCGCTNCTGA-CNTGAAGTCCCAAGTCTGGAAGAGCGGCTTCC-GGGGAGCGGCG 596
QY 856 CCGGGAAGCGGACCGCTCCTGCTCAGTGTCTCTG 890
Db 597 CCGGGAAGCGGACCGCTTGAANGCTTTNTG 631

```







[illegible]

QY 626 CGCTCAGGCTGGCTCCAGCTACATCGCCCACTTGAGGAGAGATCTGGCTACGACAAAT 685  
Db 181 CGCTCAGGCTGGCTCCAGCTACATCGCCCACTTGAGGAGAGATCTGGCTACGACAAAT 240

## RESULT 13

US-10-247-671-97  
; Sequence 97, Application US/10247671  
; Publication No. US20030194721A1  
; GENERAL INFORMATION:  
; APPLICANT: Mixita, Thomas  
; APPLICANT: Shiffman, Dov  
; APPLICANT: Porter, Gordon, J.  
; APPLICANT: Kaser, Matthew R.  
; TITLE OF INVENTION: GENES EXPRESSED IN TREATED FOAM CELLS  
; FILE REFERENCE: PA-0050 US  
; CURRENT APPLICATION NUMBER: US/10/247,671  
; CURRENT FILING DATE: 2002-09-18  
; PRIOR APPLICATION NUMBER: 60/323,784  
; PRIOR FILING DATE: 2001-09-19  
; NUMBER OF SEQ ID NOS: 186  
; SOFTWARE: PERL Program  
; SEQ ID NO 97  
; LENGTH: 2177  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; OTHER INFORMATION: Incyte ID No. US20030194721A1 3282941CB1  
US-10-247-671-97

Query Match 16.6%; Score 211.6; DB 15; Length 2177;  
Best Local Similarity 68.5%; Pred. No. 3.6e-41;  
Matches 292; Conservative 0; Mismatches 134; Indels 0; Gaps 0;  
QY 377 GCACCGAGAGAGCTCAACTGCGAGAAATGGGTCTCCCGAGAGGGCCCGCGGCGCTGG 436  
Db 651 GCAGCGGGAAGGCTGCAAGAGGAAGCGGCCCGGTGTGGCTGGGGGCGCGCGCAGGTG 710  
QY 437 GCAAGAGGAGGAGGCGGCCCAAGAGAGAGCCCTCCCGGCCAAGGGCTCAGCCAGGAGGGGA 496  
Db 711 GTAGCGGCGGCTGGTGGCAAGAGAGCCCTCCCGGCCAAGGGCTCAGCCAGGAGTGA 770  
QY 497 AGCAGTTCAGCGCAACCGCGCCCAAGAGAGAGGGCCCGCGGCGCTGGAGTGTGAGCA 556  
Db 771 AGCAGTTCAGCGCAACCGCGCCCAAGAGAGAGGGCCCGCGGCGCTGGAGTGTGAGCA 830  
QY 557 AGGCTTCTCCAGACTCAAGACACCGCTGCGGTGCGCCCGGCGGCGGCGGCTCTCCA 616  
Db 831 AGGCTTCTCCAGACTCAAGACACCGCTGCGGTGCGCCCGGCGGCGGCGGCTCTCCA 890  
QY 617 AGCTGACACGCTCAGGCTGGGTTCAGTATCATCGCTCAGCTGAGGAGATCTGGGTA 676  
Db 891 AGCTGACACGCTCAGGCTGGGTTCAGTATCATCGCTCAGCTGAGGAGATCTGGGTA 950  
QY 677 ACAGCAATACGAGACGGGTACATTCACCGGTCAACCTGAGCTGGGCGCTTATGGTG 736  
Db 951 AGGACCGCTATGAGACGGGTACATTCACCGGTCAACCTGAGCTGGGCGCTTATGGTG 1010  
QY 737 CCGGGAACCGGAGTGAACCTGAAAGAGTGGTGAACCGGCGGCGCTTATGTGAACCA 796  
Db 1011 CCGGGAACCGGAGTGAACCTGAAAGAGTGGTGAACCGGCGGCGCTTATGTGAACCA 1070  
QY 797 CCGGCT 802  
Db 1071 CCGGCT 1076

## RESULT 14

US-10-084-817-120  
; Sequence 120, Application US/10084817  
; Publication No. US20030119009A1  
; GENERAL INFORMATION:  
; APPLICANT: Wyeth  
; APPLICANT: Burczynski, Michael  
; APPLICANT: Twine, Natalie  
; APPLICANT: Dörner, Andrew  
; APPLICANT: Trepicchio, William  
; TITLE OF INVENTION: Method for Monitoring Drug Activities In Vivo  
; FILE REFERENCE: AM101080 (031896-013000)  
; CURRENT APPLICATION NUMBER: US/10/775,169  
; CURRENT FILING DATE: 2004-02-11  
; NUMBER OF SEQ ID NOS: 5278

; APPLICANT: Susan Stuart  
; APPLICANT: Jed G. Nuchtern  
; APPLICANT: Sharon E. Plon  
; APPLICANT: Jason M. Shohet  
; TITLE OF INVENTION: GENES REGULATED BY MYCN ACTIVATION  
; FILE REFERENCE: PA-0046 US  
; CURRENT APPLICATION NUMBER: US/10/084,817  
; CURRENT FILING DATE: 2002-02-25  
; PRIOR APPLICATION NUMBER: 60/270,784  
; PRIOR FILING DATE: 2001-02-23  
; NUMBER OF SEQ ID NOS: 365  
; SOFTWARE: PERL Program  
; SEQ ID NO 120  
; LENGTH: 2196  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; OTHER INFORMATION: Incyte ID No. US20030119009A1 3282941CB1  
US-10-084-817-120

Query Match 16.6%; Score 211.6; DB 15; Length 2196;  
Best Local Similarity 68.5%; Pred. No. 3.6e-41;  
Matches 292; Conservative 0; Mismatches 134; Indels 0; Gaps 0;  
QY 377 GCACCGAGAGAGCTCAACTGCGAGAAATGGGTCTCCCGAGAGGGCCCGCGGCGCTGG 436  
Db 651 GCAGCGGGAAGGCTGCAAGAGGAAGCGGCCCGGTGTGGCTGGGGGCGCGCGCAGGTG 710  
QY 437 GCAAGAGGAGGAGGCGGCCCAAGAGAGAGCCCTCCCGGCCAAGGGCTCAGCCAGGAGGGGA 496  
Db 711 GTAGCGGCGGCTGGTGGCAAGAGAGCCCTCCCGGCCAAGGGCTCAGCCAGGAGTGA 770  
QY 497 AGCAGTTCAGCGCAACCGCGCCCAAGAGAGAGGGCCCGCGGCGCTGGAGTGTGAGCA 556  
Db 771 AGCAGTTCAGCGCAACCGCGCCCAAGAGAGAGGGCCCGCGGCGCTGGAGTGTGAGCA 830  
QY 557 AGGCTTCTCCAGACTCAAGACACCGCTGCGGTGCGCCCGGCGGCGGCTCTCCA 616  
Db 831 AGGCTTCTCCAGACTCAAGACACCGCTGCGGTGCGCCCGGCGGCGGCTCTCCA 890  
QY 617 AGCTGACACGCTCAGGCTGGGTTCAGTATCATCGCTCAGCTGAGGAGATCTGGGTA 676  
Db 891 AGCTGACACGCTCAGGCTGGGTTCAGTATCATCGCTCAGCTGAGGAGATCTGGGTA 950  
QY 677 ACAGCAATACGAGACGGGTACATTCACCGGTCAACCTGAGCTGGGCGCTTATGGTG 736  
Db 951 AGGACCGCTATGAGACGGGTACATTCACCGGTCAACCTGAGCTGGGCGCTTATGGTG 1010  
QY 737 CCGGGAACCGGAGTGAACCTGAAAGAGTGGTGAACCGGCGGCGCTTATGTGAACCA 796  
Db 1011 CCGGGAACCGGAGTGAACCTGAAAGAGTGGTGAACCGGCGGCGCTTATGTGAACCA 1070  
QY 797 CCGGCT 802  
Db 1071 CCGGCT 1076

## RESULT 15

US-10-775-169-201  
; Sequence 201, Application US/10775169  
; Publication No. US20040175743A1  
; GENERAL INFORMATION:  
; APPLICANT: Wyeth  
; APPLICANT: Burczynski, Michael  
; APPLICANT: Twine, Natalie  
; APPLICANT: Dörner, Andrew  
; APPLICANT: Trepicchio, William  
; TITLE OF INVENTION: Method for Monitoring Drug Activities In Vivo  
; FILE REFERENCE: AM101080 (031896-013000)  
; CURRENT APPLICATION NUMBER: US/10/775,169  
; CURRENT FILING DATE: 2004-02-11  
; NUMBER OF SEQ ID NOS: 5278

; SOFTWARE: PatentIn version 3.2

; SEQ ID NO 201

; LENGTH: 1716

; TYPE: DNA

; ORGANISM: Homo sapiens

US-10-775-169-201

Query Match 16.5%; Score 210; DB 17; Length 1716;

Best Local Similarity 68.3%; Pred. No. 7.8e-41;

Matches 291; Conservative 0; Mismatches 135; Indels 0; Gaps 0;

QY	377	GCACCGAGGAGAGCTCCAACTGGGAGATGGGTCTCCCCAGAAAGGSCCGCGGCTGG	436
Db	180	GCAGCCGGAAGGCTGCAAGAGGAAGCGCCCCCGGGGCTGGGGCGCGCGCAGGTG	239
QY	437	GCAAGAGGAGGAAGGCGCCACCAAGAGAGCCCCCTGAGCGGGGTACCCAGAGGGGA	496
Db	240	GTAGCGCGGCGGTGGTGGCAAGAGCCCTCCCGGCCAAGGGCTCAGCGCAGAGTGCA	299
QY	497	AGCAGTCCAGCGCAACGCGCCCAACGCGGAGAGCGGCGCCGATGCGAGTGCTGAGCA	556
Db	300	AGCAGTCCAGCGCAACGCGCCCAACGCGGAGAGCGGCGCCGATGCGAGTGCTGAGCA	359
QY	557	AGGCTTTCTCCAGACTCAAGACACCCCTGCTGGTGGTGGTGGTGGTGGTGGTGGTGGT	616
Db	360	AAGCTTTCTCCAGACTCAAGACACCCCTGCTGGTGGTGGTGGTGGTGGTGGTGGTGGT	419
QY	617	AGCTGACACGCTCAGGCTGGGCTCAGCTACATGCCCCACTGAGGAGATCCTGGCTA	676
Db	420	AGCTGACACGCTCAGGCTGGGCTTCCAGTTACATGCTCACCCTGCGGAGCTGTTGCAGG	479
QY	677	ACGACAAATACGAGAACGGGTACATTACCCCGTCAACCTGACGTGGCCCTTTATGGTGG	736
Db	480	AGGACCGCTATGAGAACGGCTAGCTGACCCAGTGAACCTGACATGGCCATTCGTGGTCT	539
QY	737	CCGGGAAACCGAGAGTACCTGAAAGAGTGGTACCCGAGAGCGCTTATGTGAAACA	796
Db	540	CGGGAGAGACCGGACTCTGACACCAAGAGTTTCCGAGCCCAACAGACTATGTGGAACA	599
QY	797	CGGCGT 802	
Db	600	CGGCTT 605	

Search completed: December 16, 2004, 22:04:49

Job time : 731 secs

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: December 17, 2004, 01:07:28 ; Search time 88 Seconds  
(without alignments)  
1445.809 Million cell updates/sec

Title: US-09-701-674A-23

Perfect score: 179

Sequence: 1 MGTGSLSDVEDLQEVEMLEC.....PESDLKEVVTASRLGGTAS 179

Scoring table:

OLIGO  
Xgapop 60.0 , Xgapext 60.0  
Ygapop 60.0 , Ygapext 60.0  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 824507 seqs, 355394441 residues

Word size: 1

Total number of hits satisfying chosen parameters: 1643622

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DRV=xlh  
-O=/cgn2\_1/USPTO spoel\_h/US09701674/runat\_15122004\_100531\_28340/app query.fasta\_1.327  
-DB=Issued Patents NA -QEXT=fastap -SUFFIX=oligo.xli -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi  
-LIST=45 -DOCALLIGN=200 -THR SCORE=quality -THR MIN=1 -ALIGN=15 -MODE=LOCAL  
-OUTFTW=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US09701674 @CGN 1.1 69 @runat\_15122004\_100531\_28340 -NCPU=6 -ICPU=3  
-NO WMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG  
-DSV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6  
-FGAPEXT=7 -XGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

Issued Patents NA:  
1: /cgn2\_6/prodata/1/ina/5A COMB.seq:  
2: /cgn2\_6/prodata/1/ina/5B COMB.seq:  
3: /cgn2\_6/prodata/1/ina/6A COMB.seq:  
4: /cgn2\_6/prodata/1/ina/6B COMB.seq:  
5: /cgn2\_6/prodata/1/ina/6C COMB.seq:  
6: /cgn2\_6/prodata/1/ina/backfiles1.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	36	20.1	240	4	US-09-016-434-1028
2	9	5.0	2610	4	US-09-489-039A-1390
3	8	4.5	276	4	US-09-270-767-3094
4	8	4.5	276	4	US-09-270-767-18376
5	8	4.5	633	3	US-09-328-111-493
6	8	4.5	1013	4	US-09-270-767-10281
7	8	4.5	1893	4	US-09-489-039A-5852
8	8	4.5	2595	4	US-09-489-039A-5698
9	8	4.5	4853	2	US-08-793-824-1
10	7	3.9	51	4	US-09-443-199C-469
11	7	3.9	88	3	US-09-237-712-91
12	7	3.9	151	4	US-09-513-999C-14487

13	7	3.9	158	4	US-09-513-999C-20456
14	7	3.9	176	4	US-09-313-284A-5167
15	7	3.9	216	4	US-09-702-705-735
16	7	3.9	216	4	US-09-736-457-735
17	7	3.9	216	4	US-09-614-124B-735
18	7	3.9	216	4	US-09-671-325-735
19	7	3.9	216	4	US-09-589-184-735
20	7	3.9	216	4	US-09-658-824-735
21	7	3.9	252	4	US-09-248-786A-13618
22	7	3.9	270	4	US-09-252-991A-13449
23	7	3.9	283	4	US-09-313-294A-6600
24	7	3.9	285	4	US-09-252-991A-9893
25	7	3.9	285	4	US-09-252-991A-16285
26	7	3.9	324	4	US-09-621-976-18476
27	7	3.9	367	3	US-09-085-199B-38
28	7	3.9	375	4	US-09-513-999C-1244
29	7	3.9	380	4	US-09-513-999C-13010
30	7	3.9	399	4	US-09-489-039A-2211
31	7	3.9	405	4	US-09-621-976-14307
32	7	3.9	412	4	US-09-621-976-12072
33	7	3.9	455	4	US-09-513-999C-11196
34	7	3.9	461	4	US-09-270-767-5677
35	7	3.9	461	4	US-09-270-767-20959
36	7	3.9	465	4	US-09-252-991A-15919
37	7	3.9	477	4	US-09-489-039A-1796
38	7	3.9	477	4	US-09-270-767-7639
39	7	3.9	477	4	US-09-270-767-22921
40	7	3.9	486	3	US-09-328-111-443
41	7	3.9	493	4	US-09-702-705-193
42	7	3.9	493	4	US-09-736-457-193
43	7	3.9	493	4	US-09-614-124B-193
44	7	3.9	493	4	US-09-671-325-193
45	7	3.9	493	4	US-09-589-184-193

## ALIGNMENTS

RESULT 1  
US-09-016-434-1028  
; Sequence 1028, Application US/09016434  
; Patent No. 6500938  
; GENERAL INFORMATION:  
; APPLICANT: Janice Au-Young  
; APPLICANT: Jeffrey J. Seilhamer  
; TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF SIGNALING  
; TITLE OF INVENTION: PATHWAY GENE EXPRESSION  
; NUMBER OF SEQUENCES: 1490  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.  
; STREET: 3174 PORTER DRIVE  
; CITY: PALO ALTO  
; STATE: CALIFORNIA  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/016,434  
; FILING DATE: HEREWITH  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Zeller, Karen J.  
; REGISTRATION NUMBER: 37,071  
; REFERENCE/DOCKET NUMBER: PA-0002 US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (650) 855-0555

Sequence 20456, A  
Sequence 5167, Ap  
Sequence 735, App  
Sequence 735, App  
Sequence 735, App  
Sequence 735, App  
Sequence 735, App  
Sequence 735, App  
Sequence 735, App  
Sequence 13618, A  
Sequence 13449, A  
Sequence 6600, Ap  
Sequence 9893, Ap  
Sequence 16285, A  
Sequence 18476, A  
Sequence 38, Appl  
Sequence 1244, Ap  
Sequence 13010, A  
Sequence 2211, Ap  
Sequence 14307, A  
Sequence 12072, A  
Sequence 11196, A  
Sequence 5677, Ap  
Sequence 20959, A  
Sequence 15919, A  
Sequence 1796, Ap  
Sequence 7639, Ap  
Sequence 22921, A  
Sequence 443, App  
Sequence 193, App  
Sequence 193, App  
Sequence 193, App  
Sequence 193, App  
Sequence 193, App

```

; TELEFAX: (650) 845-4166
; INFORMATION FOR SEQ ID NO: 1028:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 240 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
;   IMMEDIATE SOURCE:
;   LIBRARY: KIDNOT05
;   CLONE: 954226
US-09-016-434-1028

Alignment Scores:
Pred. No.: 4,74e-29 Length: 240
Score: 36.00 Matches: 36
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 20.11% Indels: 0
DB: 4 Gaps: 0

US-09-701-674A-23 (1-179) x US-09-016-434-1028 (1-240)

Qy 90 AlaArgMetArgValLeuSerLysAlaPheSerArgLeuLysThrLeuProTrpVal 109
Db 90 GCCCGCATCGAGTGTGAGCAAGGCTTCTCCAGACTCAGACCCCTGCCCTGGGTG 149

Qy 110 ProProAspThrLysLeuSerLysLeuAspThrLeuArgLeuAlaSer 125
Db 150 CCCCCGACCAAGCTCTCCAAGCTGGACACGCTCAGGCTGGCGTCC 197

RESULT 2
US-09-489-039A-1390/c
; Sequence 1390, Application US/09489039A
; Patent No. 6810836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; FILE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 1390
; LENGTH: 2610
; TYPE: DNA
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-1390

Alignment Scores:
Pred. No.: 10 Length: 2610
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 5.03% Indels: 0
DB: 4 Gaps: 0

US-09-701-674A-23 (1-179) x US-09-489-039A-1390 (1-2610)

Qy 52 GlyArgGlyLeuGlyLysArgArg 60
Db 1239 GGTCCGGTGGCTTGGTAAACGGCGT 1213

RESULT 3
US-09-270-767-3094/c
; Sequence 3094, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767

; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 3094
; LENGTH: 276
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-3094

Alignment Scores:
Pred. No.: 12.9 Length: 276
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 4.47% Indels: 0
DB: 4 Gaps: 0

US-09-701-674A-23 (1-179) x US-09-270-767-3094 (1-276)

Qy 87 ArgGluArgAlaArgMetArgVal 94
Db 261 AGAGAGAGACGAAGATGCCGTG 238

RESULT 4
US-09-767-18376/c
; Sequence 18376, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 18376
; LENGTH: 276
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-18376

Alignment Scores:
Pred. No.: 12.9 Length: 276
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 4.47% Indels: 0
DB: 4 Gaps: 0

US-09-701-674A-23 (1-179) x US-09-270-767-18376 (1-276)

Qy 87 ArgGluArgAlaArgMetArgVal 94
Db 261 AGAGAGAGACGAAGATGCCGTG 238

RESULT 5
US-09-328-111-493/c
; Sequence 493, Application US/09328111
; Patent No. 6262333
; GENERAL INFORMATION:
; APPLICANT: Endege, Wilson O.
; APPLICANT: Steinmann, Kathleen E.
; APPLICANT: Astle, Jon H.
; APPLICANT: Burgess, Christopher C.
; APPLICANT: Bushnell, Steven E.
; APPLICANT: Carroll III, Eddie
; APPLICANT: Catino, Theodore J.
; APPLICANT: Derti, Adnan
; APPLICANT: Ford, Donna M.
; APPLICANT: Lewis, Marcia E.
; APPLICANT: Monahan, John E.
; APPLICANT: Schlegel, Robert
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION

```

```

; TITLE OF INVENTION: PRODUCTS
; FILE REFERENCE: CCD-257 (US)
; CURRENT APPLICATION NUMBER: US/09/328,111
; PRIOR FILING DATE: 1999-06-08
; EARLIER APPLICATION NUMBER: US 60/088,801
; EARLIER FILING DATE: 1998-06-10
; NUMBER OF SEQ ID NOS: 850
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 493
; LENGTH: 633
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)...(633)
; OTHER INFORMATION: n = A,T,C or G
US-09-328-111-493

```

```

Alignment Scores:
Pred. No.: 28.7 Length: 633
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 4.47% Indels: 0
DB: 3 Gaps: 0

```

US-09-701-674A-23 (1-179) x US-09-328-111-493 (1-633)

```

QY 105 ThrLeuProTrrpValProPasp 112
DB 470 ACTTTACCATGGTTCCTCCCGAC 447

```

#### RESULT 6

```

; Sequence 10281, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10281
; LENGTH: 1013
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-10281

```

```

Alignment Scores:
Pred. No.: 45.2 Length: 1013
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 4.47% Indels: 0
DB: 4 Gaps: 0

```

US-09-701-674A-23 (1-179) x US-09-270-767-10281 (1-1013)

```

QY 119 AspThrLeuArgLeuAlaSerSer 126
DB 706 GACACTCTGAGGCTTGCCTCCCTCC 583

```

#### RESULT 7

```

; Sequence 5852, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; FILE REFERENCE: 2709.2004001

```

```

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; FILE REFERENCE: 2709.2004001

```

```

; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 5852
; LENGTH: 1893
; TYPE: DNA
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-5852

```

```

Alignment Scores:
Pred. No.: 82.3 Length: 1893
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 4.47% Indels: 0
DB: 4 Gaps: 0

```

US-09-701-674A-23 (1-179) x US-09-489-039A-5852 (1-1893)

```

QY 122 ArgLeuAlaSerSerTyrIleAla 129
DB 1433 CGGTGGCGTCGTCATACATCGCC 1410

```

#### RESULT 8

```

; Sequence 5698, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 5698
; LENGTH: 2595
; TYPE: DNA
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-5698

```

```

Alignment Scores:
Pred. No.: 112 Length: 2595
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 4.47% Indels: 0
DB: 4 Gaps: 0

```

US-09-701-674A-23 (1-179) x US-09-489-039A-5698 (1-2595)

```

QY 122 ArgLeuAlaSerSerTyrIleAla 129
DB 563 CGGTGGCGTCGTCATACATCGCC 586

```

#### RESULT 9

```

; Sequence 1, Application US/08793824
; Patent No. 5981838
; GENERAL INFORMATION:
; APPLICANT: Simpson, Christine Lynn
; APPLICANT: Giffard, Philip Morrison
; APPLICANT: Jacques, Nicholas Anthony
; TITLE OF INVENTION: Genetic Manipulation of Plants to
; TITLE OF INVENTION: Increase Stored Carbohydrates
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Griffith Hack & Co
; STREET: Level 8, 168 Walker Street

```



```

; CITY: No. 5981838th Sydney
; STATE: New South Wales
; COUNTRY: Australia
; ZIP: 2060
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/793,824
; FILING DATE:
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: AU PM7643
; FILING DATE: 24-AUG-1994
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 61 2 9957 5944
; TELEFAX: 61 2 957 6288
; TELEX: 26547
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4853 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Streptococcus salivarius
; US-08-793-824-1

Alignment Scores:
Pred. No.: 204 Length: 4853
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 4.47% Indels: 0
DB: 2 Gaps: 0

US-09-701-674A-23 (1-179) x US-08-793-824-1 (1-4853)

Qy 165 LysGluValValThrAlaSerArg 172
Db 3241 AAAGAAGTGTACAGCTTCGT 3264

RESULT 10
US-09-443-199C-469/C
; Sequence 469, Application US/09443199C
; Patent No. 6670464
; GENERAL INFORMATION:
; APPLICANT: Shimkets, Richard A.
; TITLE OF INVENTION: Nucleic Acids Containing Single Nucleotide
; FILE REFERENCE: 15966-534A
; CURRENT APPLICATION NUMBER: US/09/443,199C
; CURRENT FILING DATE: 1999-11-16
; PRIOR APPLICATION NUMBER: 60/109,024
; PRIOR FILING DATE: 1998-11-17
; NUMBER OF SEQ ID NOS: 1272
; SOFTWARE: CuraGen Patent Formatter Version 0.9
; SEQ ID NO 469
; LENGTH: 51
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (26)...(0)
; OTHER INFORMATION: 1 of 2 allelic variants (470 is other entry)
; NAME/KEY: misc_feature
; LOCATION: (0)...(0)

```

```

; OTHER INFORMATION: Accession number cg32177197
US-09-443-199C-469

Alignment Scores:
Pred. No.: 28.6 Length: 51
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.91% Indels: 0
DB: 4 Gaps: 0

US-09-701-674A-23 (1-179) x US-09-443-199C-469 (1-51)

Qy 169 ThrAlaSerArgLeuCysGly 175
Db 39 ACGCCTCCAGATTATGTGGA 19

RESULT 11
US-09-237-712-91
; Sequence 91, Application US/09237712
; Patent No. 6180391
; GENERAL INFORMATION:
; APPLICANT: BROWN, WILLIAM C.
; TITLE OF INVENTION: HIGHLY EFFICIENT CONTROLLED EXPRESSION OF EXOGENOUS
; FILE REFERENCE: A-518
; CURRENT APPLICATION NUMBER: US/09/237,712
; CURRENT FILING DATE: 1999-01-26
; EARLIER APPLICATION NUMBER: 60/072,794
; EARLIER FILING DATE: 1998-01-28
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 91
; LENGTH: 88
; TYPE: DNA
; ORGANISM: oligonucleotide
US-09-237-712-91

Alignment Scores:
Pred. No.: 48.3 Length: 88
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.91% Indels: 0
DB: 3 Gaps: 0

US-09-701-674A-23 (1-179) x US-09-237-712-91 (1-88)

Qy 84 AlaAsnAlaArgGluArgAla 90
Db 43 GCTAACCGCAGAGAGGGCA 63

RESULT 12
US-09-513-999C-14487/C
; Sequence 14487, Application US/09513999C
; Patent No. 6783961
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Duclert, A.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; Patent No. 6783961
; FILE REFERENCE: 59 US2,REG
; CURRENT APPLICATION NUMBER: US/09/513,999C
; CURRENT FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/122,487
; PRIOR FILING DATE: 1999-02-26
; NUMBER OF SEQ ID NOS: 36681
; SOFTWARE: Patent.pm
; SEQ ID NO 14487
; LENGTH: 151
; TYPE: DNA
; ORGANISM: Homo sapiens

```

```
US-09-513-999C-14487
Alignment Scores:
Pred. No.: 81.2 Length: 151
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.91% Indels: 0
DB: 4 Gaps: 0

US-09-701-674A-23 (1-179) x US-09-513-999C-14487 (1-151)
QY 54 GYGLYLeuGLYLYsArg 60
DB 55 GGTGGACTGGCAAGAGACGA 35

RESULT 13
US-09-513-999C-20456
; Sequence 20456, Application US/09513999C
; Patent No. 6783961
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Duclert, A.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; Patent No. 6783961
; FILE REFERENCE: 59.US2.REG
; CURRENT APPLICATION NUMBER: US/09/513,999C
; CURRENT FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/122,487
; PRIOR FILING DATE: 1999-02-26
; NUMBER OF SEQ ID NOS: 36681
; SOFTWARE: Patent.pm
; SEQ ID NO 20456
; LENGTH: 158
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-513-999C-20456

Alignment Scores:
Pred. No.: 84.8 Length: 158
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.91% Indels: 0
DB: 4 Gaps: 0

US-09-701-674A-23 (1-179) x US-09-513-999C-20456 (1-158)
QY 85 AsnAlaArgGluArgAlaArg 91
DB 63 AACGGCGTGAGAGAGACCGGA 83

RESULT 14
US-09-313-294A-5167/c
; Sequence 5167, Application US/09313294A
; Patent No. 8476212
; GENERAL INFORMATION:
; APPLICANT: Lalguadi, Raghunath V.
; APPLICANT: Ito, Laura Y.
; APPLICANT: Sherman, Bradley K.
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN EAR
; FILE REFERENCE: PL-0017 US
; CURRENT APPLICATION NUMBER: US/09/313,294A
; CURRENT FILING DATE: 1999-05-14
; NUMBER OF SEQ ID NOS: 7600
; SOFTWARE: PERL Program
; SEQ ID NO 5167
; LENGTH: 176
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: misc_feature

OTHER INFORMATION: Incyte ID No. 6476212 700349712H1
; NAME/KEY: unsure
; LOCATION: 81
; OTHER INFORMATION: a, t, c, g, or other
US-09-313-294A-5167

Alignment Scores:
Pred. No.: 94.1 Length: 176
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.91% Indels: 0
DB: 4 Gaps: 0

US-09-701-674A-23 (1-179) x US-09-313-294A-5167 (1-176)
QY 67 SerProLeuSerGlyValSer 73
DB 36 TCACCGCTCTCCGGCGTCTCC 16

RESULT 15
US-09-702-705-735
; Sequence 735, Application US/09702705
; Patent No. 6504810
; GENERAL INFORMATION:
; APPLICANT: Wang, Tongtong
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Lodes, Michael A.
; APPLICANT: Fanger, Gary
; APPLICANT: Vedwick, Tom
; APPLICANT: Carter, Darrick
; APPLICANT: Retter, Marc
; APPLICANT: Mannion, Jane
; APPLICANT: Fan, Liqun
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.478C14
; CURRENT APPLICATION NUMBER: US/09/702,705
; CURRENT FILING DATE: 2000-10-30
; NUMBER OF SEQ ID NOS: 1833
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 735
; LENGTH: 216
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-702-705-735

Alignment Scores:
Pred. No.: 115 Length: 216
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.91% Indels: 0
DB: 4 Gaps: 0

US-09-701-674A-23 (1-179) x US-09-702-705-735 (1-216)
QY 29 LysGluPheValThrSerAsn 35
DB 194 AAGGAGTTTGATACATCTAAT 214

Search completed: December 17, 2004, 03:14:57
Job time : 92 secs
```

This Page Blank (uspto)

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model  
Run on: December 17, 2004, 01:22:13 ; Search time 488 Seconds  
(without alignments)  
2025.061 Million cell updates/sec

Title: US-09-701-674A-23  
Perfect score: 179  
Sequence: 1 MTGSLSDVEDLQEVEMLEC.....PSDJKVVYASRLCGTTAS 179

Scoring table: OLIGO  
Xgapop 60.0, Xgapext 60.0  
Ygapop 60.0, Ygapext 60.0  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 4093002 seqs, 2760418825 residues

Word size: 1  
Total number of hits satisfying chosen parameters: 8176338

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:  
-MODEL=frame+ p2n.model -DEV=xlh  
-O=/cgn2\_1/USPTO\_spool\_h/US09701674/runat\_15122004\_100533\_28409/app\_query.fasta\_1.327  
-DS=Published Applications NA -QMT=fastap -SUFFIX=Oligo.rnpb -MINMATCH=0.1  
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo  
-TRANS=human40.cdi -LIST=45 -DOCLALIGN=200 -THR SCORE=quality -THR MIN=1  
-ALIGN=15 -MODE=LOCAL -CUTFMF=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0  
-MAXLEN=200000000 -USRP=US09701674 @CGN 1.1 480 @runat\_15122004\_100533\_28409  
-NCPUS=6 -ICPU=3 -NO WMAP -LARGEQUERY -NEG SCORES=0 -WAIT -NSPBLOCK=100  
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60  
-XGAPOP=6 -XGAPEXT=7 -XGAPOP=60 -XGAPEXT=60 -DELOP=6 -DELEXT=7

Database : Published Applications NA:  
1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq.\*  
2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq.\*  
3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq.\*  
4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq.\*  
5: /cgn2\_6/ptodata/1/pubpna/PCTUS\_PUBCOMB.seq.\*  
6: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq.\*  
7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq.\*  
8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq.\*  
9: /cgn2\_6/ptodata/1/pubpna/US09A\_PUBCOMB.seq.\*  
10: /cgn2\_6/ptodata/1/pubpna/US09B\_PUBCOMB.seq.\*  
11: /cgn2\_6/ptodata/1/pubpna/US09C\_PUBCOMB.seq.\*  
12: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq.\*  
13: /cgn2\_6/ptodata/1/pubpna/US10A\_PUBCOMB.seq.\*  
14: /cgn2\_6/ptodata/1/pubpna/US10B\_PUBCOMB.seq.\*  
15: /cgn2\_6/ptodata/1/pubpna/US10C\_PUBCOMB.seq.\*  
16: /cgn2\_6/ptodata/1/pubpna/US10D\_PUBCOMB.seq.\*  
17: /cgn2\_6/ptodata/1/pubpna/US10E\_PUBCOMB.seq.\*  
18: /cgn2\_6/ptodata/1/pubpna/US10F\_PUBCOMB.seq.\*  
19: /cgn2\_6/ptodata/1/pubpna/US11\_NEW\_PUB.seq.\*  
20: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq.\*  
21: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	94	52.5	482	10	US-09-918-995-9069	Sequence 9069, App
2	61	34.1	697	9	US-09-954-456-829	Sequence 829, App
3	61	34.1	697	9	US-09-954-456-1245	Sequence 1245, App
4	61	34.1	697	10	US-09-960-706-1028	Sequence 1028, App
5	61	34.1	697	10	US-09-873-319-679	Sequence 679, App
6	49	27.4	632	16	US-10-264-049-1307	Sequence 1307, App
7	36	20.1	240	16	US-10-305-720-1028	Sequence 1028, App
8	28	15.6	493	10	US-09-918-995-1181	Sequence 1181, A
9	28	15.6	1716	17	US-10-775-169-201	Sequence 201, App
10	28	15.6	1939	17	US-10-755-889-365	Sequence 365, App
11	28	15.6	2177	15	US-10-247-671-97	Sequence 97, App
12	28	15.6	2196	15	US-10-084-817-120	Sequence 120, App
13	28	15.6	2382	15	US-10-240-965-83	Sequence 83, App
14	18	10.1	65	10	US-09-908-975-134	Sequence 134, App
15	9	5.0	730	9	US-09-939-825-11	Sequence 11, App
16	9	5.0	1140	17	US-10-450-826-124	Sequence 124, App
17	9	5.0	6288	15	US-10-251-186-23	Sequence 23, App
18	9	5.0	6288	15	US-10-291-172-171	Sequence 171, App
19	9	5.0	6288	16	US-10-221-278-171	Sequence 171, App
20	8	4.5	201	17	US-10-741-601-15181	Sequence 15181, A
21	8	4.5	275	11	US-09-955-216-81	Sequence 81, App
22	8	4.5	302	9	US-09-983-965-5366	Sequence 5366, App
23	8	4.5	331	16	US-10-424-599-137725	Sequence 137725, A
24	8	4.5	369	17	US-10-437-963-76590	Sequence 76590, A
25	8	4.5	405	9	US-09-983-965-5418	Sequence 5418, App
26	8	4.5	448	18	US-10-425-115-60013	Sequence 60013, A
27	8	4.5	453	16	US-10-062-674-910	Sequence 910, App
28	8	4.5	518	18	US-10-363-345A-38647	Sequence 38647, A
29	8	4.5	518	18	US-10-363-345A-38648	Sequence 38648, A
30	8	4.5	538	16	US-10-424-599-79891	Sequence 79891, A
31	8	4.5	558	10	US-09-918-995-27048	Sequence 27048, A
32	8	4.5	565	17	US-10-437-963-37962	Sequence 37962, A
33	8	4.5	612	16	US-10-424-599-42408	Sequence 42408, A
34	8	4.5	612	16	US-10-425-114-20037	Sequence 20037, A
35	8	4.5	633	9	US-09-879-536-493	Sequence 493, App
36	8	4.5	683	13	US-10-027-632-160543	Sequence 160543, A
37	8	4.5	683	15	US-10-027-632-160543	Sequence 160543, A
38	8	4.5	705	16	US-10-425-114-33618	Sequence 33618, A
39	8	4.5	735	13	US-10-027-632-21784	Sequence 21784, A
40	8	4.5	735	15	US-10-027-632-21784	Sequence 21784, A
41	8	4.5	749	18	US-10-425-115-118396	Sequence 118396, A
42	8	4.5	839	18	US-10-425-115-120281	Sequence 120281, A
43	8	4.5	840	18	US-10-425-115-85094	Sequence 85094, A
44	8	4.5	926	17	US-10-437-963-95779	Sequence 95779, A
45	8	4.5	1008	16	US-10-424-599-98612	Sequence 98612, A

# ALIGNMENTS

RESULT 1  
US-09-918-995-9069  
; Sequence 9069, Application US/09918995  
; Publication No. US20030073623A1  
; GENERAL INFORMATION:  
; APPLICANT: Hvsseq, Inc.  
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED  
; FILE OF INVENTION: FROM VARIOUS cDNA LIBRARIES  
; FILE REFERENCE: 20411-756  
; CURRENT APPLICATION NUMBER: US/09/918,995  
; CURRENT FILING DATE: 2001-07-30  
; PRIOR APPLICATION NUMBER: US/09/235,076  
; PRIOR FILING DATE: 1999-01-20  
; NUMBER OF SEQ ID NOS: 38054  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 9069  
; LENGTH: 482  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature



Score: 61.00 Matches: 61  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 34.08% Indels: 0  
DB: 9 Gaps: 0

US-09-701-674A-23 (1-179) x US-09-954-456-1245 (1-697)

Qy 119 AspThrLeuArgLeuAlaSerSerTyrrIleAlaHisLeuArgGlnIleLeuAlaAsnAsp 138  
Db 3 GACACGCTCAGGCTGGCGTCCAGCTACATCGCCACTTGAGGCAGATCTGGCTAACGAC 62  
Qy 139 LysTyrrGluAsnGlyTyrrIleHisProValAsnLeuThrTrpProPheMetValAlaGly 158  
Db 63 AAATACGAGAACGGGTACATTCACCGGTCACCTGACGTGGCCCTTTATGGTGGCGGG 122  
Qy 159 LysProGluSerAspLeuLysGluValValThrAlaSerArgLeuCysGlyThrThrAla 178  
Db 123 AAACCCGAGAGTGACCTGAAAGAGTGGTGACCGCGAGCGCGCTTATGTGGAAACCCGCG 182  
Qy 179 Ser 179  
Db 183 TCC 185

#### RESULT 4

US-09-960-706-1028  
; Sequence 1028, Application US/09960706  
; Publication No. US20030134280A1  
; GENERAL INFORMATION:  
; APPLICANT: Munger, William E.  
; TITLE OF INVENTION: Identifying Drugs for and Diagnosis of Benign Prostatic Hyperplasia  
; FILE REFERENCE: 44921-5029-01US  
; CURRENT APPLICATION NUMBER: US/09/960,706  
; PRIOR FILING DATE: 2001-09-24  
; PRIOR APPLICATION NUMBER: 60/223,323  
; PRIOR FILING DATE: 2000-08-07  
; PRIOR APPLICATION NUMBER: 09/873,319  
; PRIOR FILING DATE: 2001-06-05  
; NUMBER OF SEQ ID NOS: 1124  
; SOFTWARE: Patent In Ver. 2.1  
; SEQ ID NO 1028  
; LENGTH: 697  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: Genbank Accession No. US20030134280A1 W73859  
; NAME/KEY: unsure  
; LOCATION: (1)..(697)  
; OTHER INFORMATION: n = a or c or g or t

US-09-960-706-1028

Alignment Scores:  
Pred. No.: 3,35e-54 Length: 697  
Score: 61.00 Matches: 61  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 34.08% Indels: 0  
DB: 10 Gaps: 0

US-09-701-674A-23 (1-179) x US-09-960-706-1028 (1-697)

Qy 119 AspThrLeuArgLeuAlaSerSerTyrrIleAlaHisLeuArgGlnIleLeuAlaAsnAsp 138  
Db 3 GACACGCTCAGGCTGGCGTCCAGCTACATCGCCACTTGAGGCAGATCTGGCTAACGAC 62  
Qy 139 LysTyrrGluAsnGlyTyrrIleHisProValAsnLeuThrTrpProPheMetValAlaGly 158  
Db 63 AAATACGAGAACGGGTACATTCACCGGTCACCTGACGTGGCCCTTTATGGTGGCGGG 122  
Qy 159 LysProGluSerAspLeuLysGluValValThrAlaSerArgLeuCysGlyThrThrAla 178  
Db 123 AAACCCGAGAGTGACCTGAAAGAGTGGTGACCGCGAGCGCGCTTATGTGGAAACCCGCG 182

Qy 179 Ser 179  
Db 183 TCC 185

#### RESULT 5

US-09-873-319-679  
; Sequence 679, Application US/09873319A  
; Publication No. US20030134324A1  
; GENERAL INFORMATION:  
; APPLICANT: Munger, William E.  
; APPLICANT: Kulkarni, Prakash  
; APPLICANT: Getzenberg, Robert H.  
; APPLICANT: Waga, Iwao  
; APPLICANT: Yamamoto, Jun  
; TITLE OF INVENTION: Identifying Drugs for and Diagnosis of Benign Prostatic Hyperplasia Using Gene Expression Profiles  
; FILE REFERENCE: 44921-5029-US  
; CURRENT APPLICATION NUMBER: US/09/873,319A  
; PRIOR FILING DATE: 2001-06-05  
; PRIOR APPLICATION NUMBER: 60/223,323  
; PRIOR FILING DATE: 2000-08-07  
; NUMBER OF SEQ ID NOS: 755  
; SOFTWARE: Patent In Ver. 2.1  
; SEQ ID NO 679  
; LENGTH: 697  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: Genbank Accession No. US20030134324A1 W73859  
; NAME/KEY: unsure  
; LOCATION: (1)..(697)  
; OTHER INFORMATION: n = a or c or g or t

US-09-873-319-679

Alignment Scores:  
Pred. No.: 3,35e-54 Length: 697  
Score: 61.00 Matches: 61  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 34.08% Indels: 0  
DB: 10 Gaps: 0

US-09-701-674A-23 (1-179) x US-09-873-319-679 (1-697)

Qy 119 AspThrLeuArgLeuAlaSerSerTyrrIleAlaHisLeuArgGlnIleLeuAlaAsnAsp 138  
Db 3 GACACGCTCAGGCTGGCGTCCAGCTACATCGCCACTTGAGGCAGATCTGGCTAACGAC 62  
Qy 139 LysTyrrGluAsnGlyTyrrIleHisProValAsnLeuThrTrpProPheMetValAlaGly 158  
Db 63 AAATACGAGAACGGGTACATTCACCGGTCACCTGACGTGGCCCTTTATGGTGGCGGG 122  
Qy 159 LysProGluSerAspLeuLysGluValValThrAlaSerArgLeuCysGlyThrThrAla 178  
Db 123 AAACCCGAGAGTGACCTGAAAGAGTGGTGACCGCGAGCGCGCTTATGTGGAAACCCGCG 182  
Qy 179 Ser 179  
Db 183 TCC 185

#### RESULT 6

US-10-264-049-1307  
; Sequence 1307, Application US/10264049  
; Publication No. US20040005579A1  
; GENERAL INFORMATION:  
; APPLICANT: Birse et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PA133p1  
; CURRENT APPLICATION NUMBER: US/10/264,049  
; PRIOR FILING DATE: 2002-10-04  
; PRIOR APPLICATION NUMBER: PCT/US01/18569



GENERAL INFORMATION:  
APPLICANT: Hyseq, Inc.  
TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED  
FROM VARIOUS CDNA LIBRARIES

FILE REFERENCE: 20411-756  
CURRENT FILING DATE: 2001-07-30  
PRIOR FILING DATE: 1999-01-20  
NUMBER OF SEQ ID NOS: 38054  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 11181

LENGTH: 493  
TYPE: DNA  
ORGANISM: Homo sapiens  
NAME/KEY: misc feature  
LOCATION: (1)-(493)  
OTHER INFORMATION: n = A, T, C or G

US-09-918-995-11181  
Alignment Scores:  
Pred. No.: 1,01e-19 Length: 493  
Score: 28.00 Matches: 28  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 15.64% Indels: 0  
DB: 10 Gaps: 0

US-09-701-674A-23 (1-179) x US-09-918-995-11181 (1-493)

QY 152 TtpPrpPheMetValIaGlyLysProGluSerAspLeuLysCluValValThralaser 171  
Db 102 TGCCCTTTATGTGCGCGGAAACCCGAGGTGACCTGAAAGAGTGTGACCGGAGC 161  
QY 172 ArgLeuCysGlyThrThrAlaser 179  
Db 162 CGCTTATGTGGAACACCGGCTCC 185

## RESULT 9

US-10-775-169-201  
Sequence 201, Application US/10775169  
Publication No. US20040175743A1

GENERAL INFORMATION:  
APPLICANT: Wyeth  
APPLICANT: Burczynski, Michael  
APPLICANT: Twine, Natalie  
APPLICANT: Dörner, Andrew  
APPLICANT: Trepicchio, William  
TITLE OF INVENTION: Method for Monitoring Drug Activities In Vivo  
FILE REFERENCE: AM101080 (031896-013000)  
CURRENT FILING DATE: 2004-02-11  
CURRENT APPLICATION NUMBER: US/10/775,169  
NUMBER OF SEQ ID NOS: 5278  
SOFTWARE: PatentIn version 3.2  
SEQ ID NO 201  
LENGTH: 1716  
TYPE: DNA  
ORGANISM: Homo sapiens

US-10-775-169-201  
Alignment Scores:  
Pred. No.: 3.17e-19 Length: 1716  
Score: 28.00 Matches: 28  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 15.64% Indels: 0  
DB: 17 Gaps: 0

US-09-701-674A-23 (1-179) x US-10-775-169-201 (1-1716)

QY 106 LeuProTrpValProPheThrLysLeuSerLysLeuAspThrLeuArgLeuAlaser 125

Db 386 CTGCCCTGGTGGTCCCGACACTAAGCTCTCCAGCTGGACACGCTCCGGCTGGCTTCC 445  
QY 126 SerTyrIleAlaHisLeuArgGln 133  
Db 446 AGTTACATCGCTCACCTGGCGCAG 469

## RESULT 10

US-10-755-889-365  
Sequence 365, Application US/10755889  
Publication No. US20040171823A1  
GENERAL INFORMATION:  
APPLICANT: Bristol-Myers Squibb Company  
TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES ASSOCIATED WITH THE NF-kB  
FILE REFERENCE: D0284 NP  
CURRENT FILING DATE: 2004-01-13  
PRIOR APPLICATION NUMBER: U.S. 60/440,068  
PRIOR FILING DATE: 2003-01-14  
PRIOR APPLICATION NUMBER: U.S. 60/469,757  
PRIOR FILING DATE: 2003-05-12  
NUMBER OF SEQ ID NOS: 823  
SOFTWARE: PatentIn version 3.2  
SEQ ID NO 365  
LENGTH: 1939  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-10-755-889-365

Alignment Scores:  
Pred. No.: 3.54e-19 Length: 1939  
Score: 28.00 Matches: 28  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 15.64% Indels: 0  
DB: 17 Gaps: 0

US-09-701-674A-23 (1-179) x US-10-755-889-365 (1-1939)

QY 106 LeuProTrpValProPheThrLysLeuSerLysLeuAspThrLeuArgLeuAlaser 125  
Db 612 CTGCCCTGGTGGTCCCGACACTAAGCTCTCCAGCTGGACACGCTCCGGCTGGCTTCC 671

QY 126 SerTyrIleAlaHisLeuArgGln 133  
Db 672 AGTTACATCGCTCACCTGGCGCAG 695

## RESULT 11

US-10-247-671-97  
Sequence 97, Application US/10247671  
Publication No. US20030194721A1  
GENERAL INFORMATION:  
APPLICANT: Mikita, Thomas  
APPLICANT: Shiffman, Dov  
APPLICANT: Porter, Gordon, J.  
APPLICANT: Kaser, Matthew R.  
TITLE OF INVENTION: GENES EXPRESSED IN TREATED FOAM CELLS  
FILE REFERENCE: PA-0050 US  
CURRENT FILING DATE: 2002-09-18  
CURRENT APPLICATION NUMBER: US/10/247,671  
PRIOR FILING DATE: 2002-09-18  
PRIOR APPLICATION NUMBER: 60/323,784  
PRIOR FILING DATE: 2001-09-19  
NUMBER OF SEQ ID NOS: 186  
SOFTWARE: PERL Program  
SEQ ID NO 97  
LENGTH: 2177  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc feature  
OTHER INFORMATION: Incyte ID No. US20030194721A1 3282941CB1  
US-10-247-671-97



Alignment Scores: 3.94e-19 Length: 2177  
 Pred. No.: 28.00 Matches: 28  
 Score: 100.00% Conservative: 0  
 Percent Similarity: 100.00% Mismatches: 0  
 Best Local Similarity: 100.00% Indels: 0  
 Query Match: 15.64% Gaps: 0  
 DB:

US-09-701-674A-23 (1-179) x US-10-247-671-97 (1-2177)

Qy 106 LeuProTnpValProPaspThrLysLeuAspThrLeuArgLeuAlaSer 125  
 Db 857 CTGCCCTGGTGGCCCCGACACTAAGCTCTCCAGCTGGACACGCTCGGCTGGCTTCC 916  
 Qy 126 SerTyrIleAlaHisLeuArgGln 133  
 Db 917 AGTTACATCGCTCACCTGGCGCAG 940

## RESULT 12

US-10-084-817-120  
 ; Sequence 120, Application US/10084817  
 ; Publication No. US20030119009A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Susan Stuart  
 ; APPLICANT: Jed G. Nuchtern  
 ; APPLICANT: Sharon E. Plon  
 ; APPLICANT: Jason M. Shohet  
 ; TITLE OF INVENTION: GENES REGULATED BY MYCN ACTIVATION  
 ; FILE REFERENCE: PA-0046 US  
 ; CURRENT APPLICATION NUMBER: US/10/084,817  
 ; CURRENT FILING DATE: 2002-02-25  
 ; PRIOR APPLICATION NUMBER: 60/270,784  
 ; PRIOR FILING DATE: 2001-02-23  
 ; NUMBER OF SEQ ID NOS: 365  
 ; SOFTWARE: PERL Program  
 ; SEQ ID NO 120  
 ; LENGTH: 2196  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; NAME/KEY: misc feature  
 ; OTHER INFORMATION: Incyte ID No. US20030119009A1 3282941CB1  
 US-10-084-817-120

Alignment Scores: 3.97e-19 Length: 2196  
 Pred. No.: 28.00 Matches: 28  
 Score: 100.00% Conservative: 0  
 Percent Similarity: 100.00% Mismatches: 0  
 Best Local Similarity: 100.00% Indels: 0  
 Query Match: 15.64% Gaps: 0  
 DB:

US-09-701-674A-23 (1-179) x US-10-084-817-120 (1-2196)

Qy 106 LeuProTnpValProPaspThrLysLeuAspThrLeuArgLeuAlaSer 125  
 Db 857 CTGCCCTGGTGGCCCCGACACTAAGCTCTCCAGCTGGACACGCTCGGCTGGCTTCC 916  
 Qy 126 SerTyrIleAlaHisLeuArgGln 133  
 Db 917 AGTTACATCGCTCACCTGGCGCAG 940

## RESULT 13

US-10-240-965-83  
 ; Sequence 83, Application US/10240965  
 ; Publication No. US20030165924A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: INCYTE GENOMICS, INC.  
 ; APPLICANT: SHIFFMAN, Dov  
 ; APPLICANT: SOMOGYI, Roland  
 ; APPLICANT: LAWN, Richard M.  
 ; APPLICANT: SEILHAMER, Jeffrey J.  
 ; APPLICANT: PORTER, Gordon J.

; APPLICANT: MIKITA, Thomas  
 ; APPLICANT: TAI, Julie  
 ; TITLE OF INVENTION: GENES EXPRESSED IN FOAM CELL DIFFERENTIATION  
 ; FILE REFERENCE: PA-0025 PCT  
 ; CURRENT APPLICATION NUMBER: US/10/240,965  
 ; CURRENT FILING DATE: 2002-10-04  
 ; PRIOR APPLICATION NUMBER: 60/195,106  
 ; PRIOR FILING DATE: 2000-04-05  
 ; NUMBER OF SEQ ID NOS: 276  
 ; SOFTWARE: PERL Program  
 ; SEQ ID NO 83  
 ; LENGTH: 2382  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; NAME/KEY: misc feature  
 ; OTHER INFORMATION: Incyte ID No. US20030165924A1 977552.1  
 US-10-240-965-83

Alignment Scores: 4.27e-19 Length: 2382  
 Pred. No.: 28.00 Matches: 28  
 Score: 100.00% Conservative: 0  
 Percent Similarity: 100.00% Mismatches: 0  
 Best Local Similarity: 100.00% Indels: 0  
 Query Match: 15.64% Gaps: 0  
 DB:

US-09-701-674A-23 (1-179) x US-10-240-965-83 (1-2382)

Qy 106 LeuProTnpValProPaspThrLysLeuAspThrLeuArgLeuAlaSer 125  
 Db 861 CTGCCCTGGTGGCCCCGACACTAAGCTCTCCAGCTGGACACGCTCGGCTGGCTTCC 920  
 Qy 126 SerTyrIleAlaHisLeuArgGln 133  
 Db 921 AGTTACATCGCTCACCTGGCGCAG 944

## RESULT 14

US-09-908-975-134  
 ; Sequence 134, Application US/09908975  
 ; Publication No. US20030165843A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: SHOSHAN, Avi  
 ; APPLICANT: WASSERMAN, Alon  
 ; APPLICANT: MINTZ, Eli  
 ; APPLICANT: MINTZ, Liat  
 ; APPLICANT: FAIGLER, Simchon  
 ; TITLE OF INVENTION: OLIGONUCLEOTIDE LIBRARY FOR DETECTING RNA TRANSCRIPTS AND SPLICING  
 ; FILE REFERENCE: 36688-0005  
 ; CURRENT APPLICATION NUMBER: US/09/908,975  
 ; CURRENT FILING DATE: 2001-07-20  
 ; PRIOR APPLICATION NUMBER: US 60/287,724  
 ; PRIOR FILING DATE: 2001-05-02  
 ; PRIOR APPLICATION NUMBER: US 60/221,607  
 ; PRIOR FILING DATE: 2000-07-28  
 ; NUMBER OF SEQ ID NOS: 32337  
 ; SOFTWARE: PatentIn version 3.0  
 ; SEQ ID NO 134  
 ; LENGTH: 65  
 ; TYPE: DNA  
 ; ORGANISM: Rattus norvegicus  
 US-09-908-975-134

Alignment Scores: 4.92e-10 Length: 65  
 Pred. No.: 18.00 Matches: 18  
 Score: 100.00% Conservative: 0  
 Percent Similarity: 100.00% Mismatches: 0  
 Best Local Similarity: 100.00% Indels: 0  
 Query Match: 10.06% Gaps: 0  
 DB:

US-09-701-674A-23 (1-179) x US-09-908-975-134 (1-65)

Qy 1 MetSerThrGlySerLeuSerAspValGluAspLeuGlnGluValGluMetLeu 18  
 Db 10 ATGTCCACTGGCTCCCTCAGCGATGTAGAAGACCTTCAAGAGGTGGAGATGCTG 63

RESULT 15

US-09-939-825-11/c  
 ; Sequence 11, Application US/09939825  
 ; Patent No. US20020151009A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Ni et al.  
 ; TITLE OF INVENTION: PT046P1  
 ; FILE REFERENCE: Human Polynucleotides, Polypeptides, and Antibodies  
 ; CURRENT APPLICATION NUMBER: US/09/939,825  
 ; CURRENT FILING DATE: 2001-08-28  
 ; PRIOR APPLICATION NUMBER: PCT/US01/05498  
 ; PRIOR FILING DATE: 2001-02-22  
 ; PRIOR APPLICATION NUMBER: 60/184,664  
 ; PRIOR FILING DATE: 2000-02-24  
 ; PRIOR APPLICATION NUMBER: 60/189,874  
 ; PRIOR FILING DATE: 2000-03-16  
 ; NUMBER OF SEQ ID NOS: 35  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 11  
 ; LENGTH: 730  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 US-09-939-825-11

Alignment Scores:  
 Pred. No.: 12.4 Length: 730  
 Score: 9.00 Matches: 9  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 5.03% Indels: 0  
 DB: 9 Gaps: 0

US-09-701-674A-23 (1-179) x US-09-939-825-11 (1-730)

Qy 54 GlyGlyLeuGlyLysArgArgLysala 62  
 Db 128 GGGGGCTTGGGAGAGAGAGAGGCA 102

Search completed: December 17, 2004, 04:20:36  
 Job time : 549 secs

**This Page Blank (uspto)**

GenCore version 5.1.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: December 17, 2004, 01:03:08 ; Search time 2869 Seconds  
(without alignments)  
2273.514 Million cell updates/sec

Title: US-09-701-674A-23

Perfect score: 179  
Sequence: 1 MTGSLSDVEDLQVEMLEC.....PESDLKVVVTASRLCGTTAS 179

Scoring table: OLIGO  
Xgapop 60.0 , Xgapext 60.0  
Ygapop 60.0 , Ygapext 60.0  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 32822875 seqs, 18219865908 residues

Word size: 1  
Total number of hits satisfying chosen parameters: 65644297

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame+pn.model -DEV=xlh  
-Q/cgn2.1/USPTO.spool\_h/US09701674/runat\_15122004\_100530\_28319/app\_query.fasta\_1.327  
-DB=EST -Qfmt=fastap -SUFFIX=oligo.rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR SCORE=QUALITY -THR MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFMT=pto  
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US09701674@cgn\_1.1\_3437@runat\_15122004\_100530\_28319 -NCPU=6 -ICPU=3  
-NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DSV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : EST.\*  
1: gb\_est1.\*  
2: gb\_est2.\*  
3: gb\_hic.\*  
4: gb\_est3.\*  
5: gb\_est4.\*  
6: gb\_est5.\*  
7: gb\_est6.\*  
8: gb\_gss1.\*  
9: gb\_gss2.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	179	100.0	782	6	CF128484 UI-HF-ETO
2	179	100.0	796	5	BX107489
3	179	100.0	1044	1	AL550251
4	179	100.0	1221	3	CR594669
5	179	100.0	1250	3	CR606057
6	179	100.0	1252	3	CR616308
7	178	99.4	1038	4	BM544256
8	177	98.9	800	6	CF127519
9	162	90.5	1099	1	AL552119

10	160	89.4	965	1	AL552380
11	153	85.5	1098	4	BM923344
12	151	84.4	757	6	CF126927
13	149	83.2	746	4	BI770998
14	138	77.1	645	6	CF128246
15	138	77.1	744	6	CF128914
16	138	77.1	785	6	CF127896
17	138	77.1	794	6	CF127034
18	136	76.0	738	6	CF130905
19	136	76.0	852	7	CF994025
20	134	74.9	895	5	CF994025
21	133	74.3	827	5	CF994025
22	133	74.3	827	5	CF994025
23	129	72.1	724	6	CF125244
24	127	70.9	1167	4	BM811192
25	125	69.8	563	1	AV605533
26	124	69.3	926	4	BI769090
27	120	67.0	1012	6	BY710826
28	120	67.0	1238	3	AK011575
29	118	65.9	526	1	AJ650722
30	117	65.4	778	7	CO564290
31	117	65.4	818	4	BI767574
32	116	64.8	712	5	EX922447
33	115	64.2	726	5	EX920031
34	112	62.6	629	6	CB447077
35	111	62.0	651	4	BI765033
36	110	61.5	589	6	CB417689
37	108	60.3	605	4	BI540239
38	108	60.3	784	7	CK770983
39	108	60.3	1085	1	AL574170
40	107	59.8	717	6	CF126467
41	107	59.8	1065	5	BQ072411
42	102	57.0	1169	1	AL552084
43	100	55.9	615	2	AW920010
44	100	55.9	1024	1	AL575799
45	98	54.7	597	5	BU358809

ALIGNMENTS

RESULT 1  
CF128484

LOCUS UI-HF-ETO-awj-c-15-0-UI.r1 NIH.MGC\_214 Homo sapiens cDNA clone  
DEFINITION IMAGE:30557534 5', mRNA sequence.  
ACCESSION CF128484 782 bp mRNA linear EST 05-AUG-2003  
VERSION CF128484.1 GI:33207776  
KEYWORDS EST.

SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 782)  
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.  
TITLE Normalization and subtraction: two approaches to facilitate gene discovery  
JOURNAL Genome Res. 6 (9), 791-806 (1996)  
MEDLINE 97044477  
PUBMED 889548  
COMMENT Contact: Soares, MB  
Coordinated Laboratory for Computational Genomics  
University of Iowa  
375 Newton Road , 4156 MEBRF, Iowa City, IA 52242, USA  
Tel: 319 335 8250  
Fax: 319 335 9565  
Email: bento-soares@uiowa.edu  
Tissue Procurement: Mary Hendrix  
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa  
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
Clone Distribution: Distribution information can be found at  
http://genome.uiowa.edu/distribution/humanfl.html  
The following repetitive elements were found in this cDNA

sequence: 1-81, >(GA)n\$imple\_repeat (matched complement)

Seq primer: pYX-5

# FEATURES

source

Location/Qualifiers

1..782

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone\_lib="DH10B (T1 phage resistant)"

/lab\_host="DH10B (T1 phage resistant)"

/notes="Organ: Lung; Vector: pYX-Asc; Site 1: EcoR I; Site 2: Not I; The library was constructed according to Bonaudo, Lennon and Soares, Genome Research, 6:791-806, 1996. Denatured RNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with oligo-dT primer containing a Not I site. Double strand cDNA was size selected according to mRNA size fraction, ligated with EcoR I adaptor, digested with Not I and then cloned directionally into pYX-Asc vector. The library tag sequence located between the Not I site and the polyA tail is GATAGGCGCA. Tissue was provided by Mary Hendrix."

## ORIGIN

Alignment Scores:  
Pred. No.: 3,01e-169 Length: 782  
Score: 179.00 Matches: 179  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 6 Gaps: 0

US-09-701-674A-23 (1-179) x CF128484 (1-782)

QY 1 MetSerThrGlySerLeuSerAspValGluAspLeuGlnGluValGluMetLeuGluCys 20  
DB 86 ATGTCACCGGCTCCCTCAGCGATGTGGAGGACCTTCAGAGGTGGAGATGTGGAATGT 145  
QY 21 AspGlyLeuLysMetAspSerAsnLysGluPheValThrSerAsnGluSerThrGluGlu 40  
DB 146 GAGCGGTTGAAATGATTCGAACAGGAAATTTGTACTTCCACGAGCAGCACCAGGAG 205  
QY 41 SerSerAsnGlySerProGlnLysGlyValGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 60  
DB 206 AGTCTCAACTCGAGAAATGGGTCTCCCAAGAGGCGCGCGCGCTGGCCAGAGGAGG 265  
QY 61 LysAlaProThrLysLysSerProLeuSerGlyValSerGlnGluGlyLysGlnValGln 80  
DB 266 AAGCGGCCACCAAGAAGAGCCCTCAGCGGGTCCAGCCAGGAGGAGCAGGTCCAG 325  
QY 81 ArgAsnAlaAlaAsnAlaArgGluArgAlaArgMetArgValLeuSerLysAlaPheSer 100  
DB 326 CGCAACGCGCCCAACCGCGAGAGCGGCGCCGATCGCGAGTGTGAGCAAGGCCCTTCCTCC 385  
QY 101 ArgLeuLysThrThrLeuProThrValProProAspThrLysLeuSerLysLeuAspThr 120  
DB 386 AGACTCAAGACCACTTCGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 445  
QY 121 LeuArgLeuAlaSerSerThrLeuAlaHisLeuArgGlnLeuAlaAsnAspLysThr 140  
DB 446 CTCAGGCTGGCGTCCAGCTACATCGCCCACTTGGAGCAGATCTCTGGTCAACGACAAATAC 505  
QY 141 GluAsnGlyThrIleHisProValAsnLeuThrTrpPheMetValAlaGlyLysPro 160  
DB 506 GAGACGGGTACATTCACCGGTCAACCTGAGCTGGTGGTGGTGGTGGTGGTGGTGGTGGT 565  
QY 161 GluSerAspLeuLysGluValValThrAlaSerArgLeuGlyThrThrAlaSer 179  
DB 566 GAGAGTGACCTGAAGAAGTGGTGTACCGGAGCGGCTTATGTGGAACCAACCGCGTCC 622

## RESULT 2

BX107489

LOCUS

796 bp

mRNA

linear

EST 06-FEB-2003

## DEFINITION

BX107489 Soares placenta NB2HP Homo sapiens cDNA clone  
IMAGE998N24234 ; IMAGE:151511, mRNA sequence.

## ACCESSION

BX107489

## VERSION

BX107489.1

## KEYWORDS

EST.

## SOURCE

Homo sapiens (human)

## REFERENCE

1 (bases 1 to 796)

## AUTHORS

Eberl,L., Heil,O., Hennig,S., Neubert,P., Partsch,E., Peters,M., Radelof,U., Schneider,D. and Korn,B.

## TITLE

Human Unigeneset - RZPD3

## JOURNAL

Unpublished (2003)

## COMMENT

Contact: Ina Rolfs

RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH

Im Neuenheimer Feld 580, D-69120 Heidelberg, Germany

RZPD; IMAGE998N24234.

RZPDLIB; I.M.A.G.E. cDNA Clone Collection;

Human Unigeneset - RZPD3 (RZPDLIB No.972).

http://www.rzpd.de/CloneCards/cgi-

bin/showLib.pl.cgi/response?libNo=972 Contact: Ina Rolfs

RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH

Heubnerweg 6, D-14059 Berlin, Germany

Tel: +49 30 32839 101

Fax: +49 30 32839 111

www.rzpd.de

This clone is available royalty-free from RZPD;

contact RZPD (clone@rzpd.de) for further information. Seq primer:

M13r, Primer sequence: TTTCACAGGAACAGCTATGAC.

## FEATURES

Location/Qualifiers

1..796

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="IMAGE998N24234 ; IMAGE:151511"

/sex="Female"

/dev\_stage="Placenta obtained at birth (full term)"

/lab\_host="DH10B (ampicillin resistant)"

/clone\_lib="Soares placenta NB2HP"

/notes="Organ: placenta; Vector: pYX-Asc (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer (5' AACTGGAAGATTTCGGCGGCGGAGGATTTTTTTTTTTT 3', double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pYX-Asc vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Fatima Bonaudo. "

## ORIGIN

Alignment Scores:  
Pred. No.: 3,06e-169 Length: 796  
Score: 179.00 Matches: 179  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 5 Gaps: 0

US-09-701-674A-23 (1-179) x BX107489 (1-796)

QY 1 MetSerThrGlySerLeuSerAspValGluAspLeuGlnGluValGluMetLeuGluCys 20  
DB 195 ATGTCCACCGGCTCCCTCAGCGATGTGGAGGACCTTCAGAGGTGGAGATGTGGAATGT 254  
QY 21 AspGlyLeuLysMetAspSerAsnLysGluPheValThrSerAsnGluSerThrGluGlu 40  
DB 255 GACGGGTGAAATGGAATTCGAACAGGAAATTTGTGACTTCCACGAGCAGCAGGAG 314  
QY 41 SerSerAsnGlySerProGlnLysGlyValGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 60  
DB 315 AGTCCCAACTGCCAGAAATGGGTCTCCCAAGAGGCGCGCGGCGCTCGGCAAGAGGAG 374



/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CS0D1039YJ05"  
/tissue\_type="Placenta Cot 25-normalized"  
/plasmid="PCMVSPORT\_6"

## ORIGIN

Alignment Scores:  
Pred. No.: 4,54e-169 Length: 1221  
Score: 179.00 Matches: 179  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 3 Gaps: 0

US-09-701-674A-23 (1-179) x CR594669 (1-1221)

Qy 1 MetSerThrGlySerLeuSerAspValGluAspLeuGlnGluValGluMetLeuGluCys 20  
Db ATGTCACCGGCTCCCTCAGCATGTGGAGGACCTTCAAGAGGTGAGATGTGGAATGT 319  
Qy 21 AspGlyLeuLysMetAspSerAsnLysGluPheValThrSerAsnGluSerThrGluGlu 40  
Db GACGGGTGAAATGATTCGAACAGGAATTTGTGACTTCCACGAGACACCGAGGAG 379  
Qy 41 SerSerAsnCysGluAsnGlySerProGlnLysGlyArgGlyGlyLeuGlyLysArg 60  
Db AGCTCCAACTGGAGAAATGGGTCTCCCAAGAGGCGCGCGCTGGGCAAGAGGAGG 439  
Qy 61 LysAlaProThrLysLysSerProLeuSerGlyValSerGlnGluGlyLysGlnValGln 80  
Db AAGCGCCCAACAGAGAGGCGGCTGAGGGGTGAGCGAGGAGGAGGAGGAGGAGG 499  
Qy 81 ArgAsnAlaAlaAsnAlaArgGluAlaArgMetArgValLeuSerLysAlaPheSer 100  
Db CGAACCGCGCAACCGGAGAGCGGCGCGCTGAGGAGTCTGAGCAAGGCTTCTCC 559  
Qy 101 ArgLeuLysThrThrLeuProTrpValProProAspThrLysLeuSerLysLeuAspThr 120  
Db AGACTCAGACCACTTGGCTGGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 619  
Qy 121 LeuArgLeuAlaSerSerTyrIleAlaHisLeuArgGlnIleLeuAlaAsnAspLysTyr 140  
Db CTGAGCTGGCGTCCAGCTACATCGCCACTTGAGGAGATCTGAGGAGTCTGAGGAG 679  
Qy 141 GluAsnGlyTyrIleHisProValAsnLeuThrTrpProPheMetValAlaGlyLysPro 160  
Db GAGAACGGGTACATTCAACCGGTCAACCTGAGTGGGCTTTATGGTGGCGGGAAACCC 739  
Qy 161 GluSerAspLeuLysGluValValThrAlaSerArgLeuCysGlyThrThrAlaSer 179  
Db GAGAGTGACCTGAAGAAGTGTGTGACCGGAGCGGCTTATGTGGAACCAACCGGCTCC 796

RESULT 5  
CR606057 1250 bp mRNA linear HTC 21-JUL-2004  
LOCUS full-length cDNA clone CS0D1039YJ05 of Placenta Cot 25-normalized  
DEFINITION of Homo sapiens (human)  
ACCESSION CR606057  
VERSION CR606057.1 GI:50486864  
KEYWORDS HIT; CDS; mRNA  
SOURCE Homo sapiens  
ORGANISM Homo sapiens (human)  
REFERENCE 1 (bases 1 to 1250)  
AUTHORS Li W.B., Gruber, C., Jessee, J. and Polayes, D.  
TITLE Full-length cDNA libraries and normalization  
JOURNAL Unpublished  
REMARK Contact : Feng Liang Email : fliang@lifetech.com URL :  
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600  
Faraday Avenue  
2 (bases 1 to 1250)

AUTHORS Genoscope.  
TITLE Direct Submission  
JOURNAL Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :  
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)  
COMMENT - Web : www.genoscope.cns.fr)  
18c strand cDNA was primed with a NotI-oligo(dT) primer. Five prime  
end enriched, double-strand cDNA was digested with Not I and cloned  
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library  
was normalized. Library was constructed by Life Technologies, a  
division of Invitrogen.

## FEATURES

source Location/Qualifiers  
1..1250  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CS0D1039YJ05"  
/tissue\_type="Placenta Cot 25-normalized"  
/plasmid="PCMVSPORT\_6"

## ORIGIN

Alignment Scores:  
Pred. No.: 4,63e-169 Length: 1250  
Score: 179.00 Matches: 179  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 3 Gaps: 0

US-09-701-674A-23 (1-179) x CR606057 (1-1250)

Qy 1 MetSerThrGlySerLeuSerAspValGluAspLeuGlnGluValGluMetLeuGluCys 20  
Db ATGTCACCGGCTCCCTCAGCATGTGGAGGACCTTCAAGAGGTGAGATGTGGAATGT 337  
Qy 21 AspGlyLeuLysMetAspSerAsnLysGluPheValThrSerAsnGluSerThrGluGlu 40  
Db GACGGGTGAAATGATTCGAACAGGAATTTGTGACTTCCAAACGAGACACCGAGGAG 397  
Qy 41 SerSerAsnCysGluAsnGlySerProGlnLysGlyArgGlyGlyLeuGlyLysArg 60  
Db AGCTCCAACTGGAGAAATGGGTCTCCCAAGAGGCGCGCGCTGGGCAAGAGGAGG 457  
Qy 61 LysAlaProThrLysLysSerProLeuSerGlyValSerGlnGluGlyLysGlnValGln 80  
Db AAGCGCGCCCAACAGAGAGCGGCTGAGCGGGTCAAGAGGAGGAGGAGGAGGAGG 517  
Qy 81 ArgAsnAlaAlaAsnAlaArgGluAlaArgMetArgValLeuSerLysAlaPheSer 100  
Db CGAACCGCGCAACCGGAGAGCGGCGCGCTGAGGAGTCTGAGCAAGGCTTCTCC 577  
Qy 101 ArgLeuLysThrThrLeuProTrpValProProAspThrLysLeuSerLysLeuAspThr 120  
Db AGACTCAGACCACTTGGCTGGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 637  
Qy 121 LeuArgLeuAlaSerSerTyrIleAlaHisLeuArgGlnIleLeuAlaAsnAspLysTyr 140  
Db CTGAGCTGGCGTCCAGCTACATCGCCACTTGAGGAGATCTGAGGAGTCTGAGGAG 697  
Qy 141 GluAsnGlyTyrIleHisProValAsnLeuThrTrpProPheMetValAlaGlyLysPro 160  
Db GAGAACGGGTACATTCAACCGGTCAACCTGAGTGGGCTTTATGGTGGCGGGAAACCC 757  
Qy 161 GluSerAspLeuLysGluValValThrAlaSerArgLeuCysGlyThrThrAlaSer 179  
Db GAGAGTGACCTGAAGAAGTGTGTGACCGGAGCGGCTTATGTGGAACCAACCGGCTCC 814

## RESULT 6

CR616308 1252 bp mRNA linear HTC 21-JUL-2004  
LOCUS full-length cDNA clone CS0D1070YK16 of Placenta Cot 25-normalized  
DEFINITION of Homo sapiens (human)  
ACCESSION CR616308  
VERSION CR616308.1 GI:50497115

```

KEYWORDS      HTC; CNSLT_CDNA.
SOURCE        Homo sapiens (human)
ORGANISM      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE     1 (bases 1 to 1252)
AUTHORS      Li, W.B., Gruber, C., Jessee, J., and Polayes, D.
TITLE        Full-length cDNA libraries and normalization
JOURNAL      Unpublished
REMARK       Contact : Feng Liang Email : fliang@lifetech.com URL :
              http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
              Faraday Avenue
              Genoscope.
REFERENCE     2 (bases 1 to 1252)
AUTHORS      Genoscope.
TITLE        Direct Submission
JOURNAL      Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :
              BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
              - Web : www.genoscope.cns.fr)
COMMENT       1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
              end enriched, double-strand cDNA was digested with Not I and cloned
              into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
              was normalized. Library was constructed by Life Technologies, a
              division of Invitrogen.
FEATURES      Location/Qualifiers
               1..1252
                /organism="Homo sapiens"
                /mol_type="mRNA"
                /db_xref="taxon:9606"
                /clone="CS01070YK16"
                /issue_type="Placenta Cot 25-normalized"
                /plasmid="pCMVSPORT_6"
ORIGIN
Alignment Scores:
Pred. No.:      4,648-169      Length:      1252
Score:          179.00         Matches:    179
Percent Similarity: 100.00%     Conservative: 0
Best Local Similarity: 100.00%  Mismatches:  0
Query Match:    100.00%        Indels:     0
DB:             3              Gaps:       0

US-09-701-674A-23 (1-179) x CR616308 (1-1252)

QY      1 MetSerThrGlySerLeuSerAspValGluAspLeuGlnGluValGluMetLeuGluCys 20
Db      272 ATGTCCACCGGCTCCCTCAGCATGTGGAGGACCTTCAAGAGGTGGAGATGTGGAATGT 331
QY      21 AspGlyLeuLysMetAspSerAsnLysGluPheValThrSerAsnGluSerThrGluGlu 40
Db      332 GACGGGTTGAAATGGATTTCGAACAAGGAATTGTGACTTCCACGAGGACCCGAGGAG 391
QY      41 SerSerAsnGluAsnGlySerProGlnLysGlyArgGlyLeuGlyLysArgArg 60
Db      392 AGCTCCAACTGGAGAAATGGGTCTCCCAAGAGGGCGCGGCTGGGCAAGAGGAGG 451
QY      61 LysAlaProThrLysLysSerProLeuSerGlyValSerGlnGluGlyLysGlnValGln 80
Db      452 AAGCGCGCCCAACAAGAAGAGCCCTCAGCGGGGTGAGCCAGGAGGGGAAGCAGGTCCAG 511
QY      81 ArgAsnAlaAlaAsnAlaArgGluArgAlaArgMetArgValLeuSerLysAlaPheSer 100
Db      512 CGCAACGCCGCCAACGCCGAGAGGGGGCCCATCGAGTCTCGAGCAGGCGCTTCCTCC 571
QY      101 ArgLeuLysThrThrLeuProTrpValProProAspThrLysLeuSerLysLeuAspThr 120
Db      572 AGACTCAAGACCAACACCTGGCTGGGTGGCCCGGACCAACCAAGCTCTCCAAGTGGACACG 631
QY      121 LeuArgLeuAlaSerSerTyrIleAlaHisLeuArgGlnIleLeuAlaAsnAspLysTyr 140
Db      632 CTCAGGCTGGCGTCCAGCTACATCGCCCATTTGAGGCAGATCCTGGCTAAACACAAATAC 691
QY      141 GluAsnGlyTyrIleHisProValAsnLeuThrTrpProPheMetValAlaGlyLysPro 160

```

---

```

Db      692 GAGAACGGGTACATTCACCCGGTCAACCTGACGTGGCCCTTTATGGTGGCGGGAACCC 751
QY      161 GluSerAspLeuLysGluValValThrAlaSerArgLeuCysGlyThrThrAlaSer 179
Db      752 GAGAGTGACCTGAAGAGAGTGGTGACCGGCGCTTATGTGGAACACCGCGCTCC 808

RESULT 7
BM544256
LOCUS      BM544256
DEFINITION AGENCOURT 6490671 NIH_MGC_125 Homo sapiens cDNA clone IMAGE:5587760
           5', mRNA sequence.
ACCESSION  BM544256
VERSION     BM544256.1 GI:18775358
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 1038)
AUTHORS     NIH-MGC http://mgi.nci.nih.gov/.
TITLE       National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL     Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
           Email: cgapbs-remail.nih.gov
           Tissue Procurement: Invitrogen
           cDNA Library Preparation: Life Technologies, Inc.
           cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
           DNA Sequencing by: Agencourt Bioscience Corporation
           Clone Distribution: MGC clone distribution information can be
           found through the I.M.A.G.E. Consortium/LNL at:
           http://image.llnl.gov
           Plate: L1AM12357 row: m column: 09
           High quality sequence stop: 774.
FEATURES      Location/Qualifiers
               1..1038
                /organism="Homo sapiens"
                /mol_type="mRNA"
                /db_xref="taxon:9606"
                /clone="IMAGE:5587760"
                /lab_host="DH10B"
                /clone_lib="NIH_MGC_125"
                /notes="Organ: ovary (pool of 3); Vector: pCMV-SPORT6;
                Site 1: EcoRV (destroyed); Site 2: NotI; RNA source pool
                of three ovaries, from females ranging in age from 38 to
                49 yo. Library is oligo-dT primed and directionally cloned
                (EcoRV site is destroyed upon cloning). Average insert
                size 2.1 kb, insert size range 1-3.5 kb. Library is
                normalized and enriched for full-length clones and was
                constructed by C. Gruber (Invitrogen). Research Genetics
                tracking code 036."
ORIGIN
Alignment Scores:
Pred. No.:      3,958-168      Length:      1038
Score:          178.00         Matches:    178
Percent Similarity: 100.00%     Conservative: 0
Best Local Similarity: 100.00%  Mismatches:  0
Query Match:    99.44%         Indels:     0
DB:             4              Gaps:       0

US-09-701-674A-23 (1-179) x BM544256 (1-1038)

QY      2 SerThrGlySerLeuSerAspValGluAspLeuGlnGluValGluMetLeuGluCysAsp 21
Db      82 TCCACCGGCTCCCTCAGCATGTGGAGGACCTTCAAGAGGTGGAGATGTGGAATGTGAC 141
QY      22 GlyLeuLysMetAspSerAsnLysGluPheValThrSerAsnGluSerThrGluGluSer 41
Db      142 GGGTTGAAATGGATTTCGAACAAGGAATTGTGACTTCCACGAGGACCCGAGGAGAGC 201
QY      42 SerAsnGluAsnGlySerProGlnLysGlyArgGlyLysArgArgLys 61
Db      202 TCCAACTGCAGAAATGGTCTCCCAAGAGGGCGCGGCTGGCAAGAGGAGGAG 261

```



gel. First strand cDNA synthesis was primed with oligo-dT primer containing a Not I site. Double strand cDNA was size selected according to mRNA size fraction, ligated with EcoR I adaptor, digested with Not I and then cloned directionally into pYX-Asc vector. The library tag sequence located between the Not I site and the polyA tail is GATAGGCCA. Tissue was provided by Mary Hendrix."

## ORIGIN

Alignment Scores: 3.15e-167 Length: 800  
 Pred. No.: 177.00 Matches: 177  
 Score: 100.00% Conservative: 0  
 Percent Similarity: 100.00% Mismatches: 0  
 Best Local Similarity: 98.88% Indels: 0  
 Query Match: 6 Gaps: 0  
 DB:

US-09-701-674a-23 (1-179) x CF127519 (1-800)

Qy 1 MetSerThrGlySerLeuSerAspValGluAspLeuGlnGluValGluMetLeuGluCys 20  
 Db 258 ATGTCCACCGGCTCCCTCAGCGATGTGGAGACCTTCAAGAGGTGGAGATGTTGGAAATGT 317  
 Qy 21 AspGlyLeuLysMetAspSerAsnLysGluPheValThrSerAsnGluSerThrGluGlu 40  
 Db 318 GACGGGTGTAATGGAATTCGAACAAGGAATTTGTGACTTCCAAAGAGACCCGAGGAG 377  
 Qy 41 SerSerAsnGlyGluAsnGlySerProGlnLysGlyArgGlyLeuGlyLysArgArg 60  
 Db 378 AGTCCCACTGCCAGATGGTCTCCCAAGAGGCGCGGCGCTGGCAAGAGGAGG 437  
 Qy 61 LysAlaProThrLysLysSerProLeuSerGlyValSerGlnGlnGlyLysGlnValGln 80  
 Db 438 AAGCGGCCCAACCAAGAGAGAGCCCTCAGCGGGTTCAGCCAGGAGGAGAGGAGTCCAG 497  
 Qy 81 ArgAsnAlaAlaAsnAlaArgGluArgAlaArgMetArgValLeuSerLysAlaPheSer 100  
 Db 498 CGCAACGCCGCCAAGCGGAGAGCGGCGCGCATCGAGTGTGAGCAAGGCTTCTCC 557  
 Qy 101 ArgLeuLysThrThrLeuProThrValProProAspThrLysLeuSerLysLeuAspThr 120  
 Db 558 AGACTCAAGACCACTCCCTGGTGGTCCCGCCGACCAAGAGCTCTCCAAGCTGGACAG 617  
 Qy 121 LeuArgLeuAlaSerSerThrAlaHisLeuArgGlnLeuAlaAsnAspLysTyr 140  
 Db 618 CTGAGGTGGCGTCCAGCTACATCGCCCATTTGAGGAGATCTCTGGCTAACGCAATAC 677  
 Qy 141 GluAsnGlyTyrIleHisProValAsnLeuThrTrpProPheMetValAlaGlyLysPro 160  
 Db 678 GAGAACGGGTACATTTCACCGGTCACCTGACGTGGCCCTTTATGTGGCGGAAACCC 737  
 Qy 161 GluSerAspLeuLysGluValValThrAlaSerArgLeuCysGlyThrThr 177  
 Db 738 GAGAGTGACCTGAAGAAGTGGTGACCGGAGCGCTTATGTGGACCAACAG 788

## RESULT 9

AL552119 1099 bp mRNA linear EST 30-MAR-2004  
 AL552119 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens CDNA  
 Clone SCODI059YH17 5-PRIME, mRNA sequence.  
 DEFINITION  
 AL552119  
 ACCESSION  
 VERSION  
 KEYWORDS  
 EST.  
 SOURCE  
 Homo sapiens (human)  
 ORGANISM  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE  
 1 (bases 1 to 1099)  
 Li, W.B., Gruber, C., Jessee, J. and Polayes, D.  
 Full-length cDNA libraries and normalization  
 Unpublished (2001)  
 JOURNAL  
 On Feb 15, 2001 this sequence version replaced gi:31273935.  
 COMMENT  
 Contact: Genoscope

Qy 62 AlaProThrLysSerProLeuSerGlyValSerGlnGluGlyLysGlnValGlnArg 81  
 Db 262 GCGCCCAACAAGAGGCGGCTGAGCGGGTCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 321  
 Qy 82 AsnAlaAlaAsnAlaArgGluArgAlaArgMetArgValLeuSerLysAlaPheSerArg 101  
 Db 322 AAGCGCGCCAAACGCGAGAGCGGGCCCGCATGCGAGTGTCTGAGCAAGGCTTCTCCAGA 381  
 Qy 102 LeuLysThrThrLeuProThrValProProAspThrLysLeuSerLysLeuAspThrLeu 121  
 Db 382 CTCAGAGCACCGCTGCGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 441  
 Qy 122 ArgLeuAlaSerSerThrIleAlaHisLeuArgGlnLeuAlaAsnAspLysTyrGlu 141  
 Db 442 AGCTGGCGTCCAGTACATCGCCCACTTGAAGGAGATCTCTGGCTAACGCAATACAG 501  
 Qy 142 AsnGlyTyrIleHisProValAsnLeuThrTrpProPheMetValAlaGlyLysProGlu 161  
 Db 502 AACCGGTACATTCACCGGCTCAACCTGACGTGGCGCTTTATGTGGCGGCAACCGGAG 561  
 Qy 162 SerAspLeuLysGluValThrAlaSerArgLeuCysGlyThrThrAlaSer 179  
 Db 562 AGTGACCTGAAGAAGTGGTGACCGGAGCGGCTTATGTGGAACCAACCGCGTCC 615

## RESULT 8

CF127519 800 bp mRNA linear EST 05-AUG-2003  
 UI-HP-ET0-awh-f-14-0-UI.r1 NIH\_MGC\_214 Homo sapiens CDNA clone  
 IMAGE:30554533 5', mRNA sequence.  
 DEFINITION  
 CF127519  
 ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 Homo sapiens (human)  
 ORGANISM  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE  
 1 (bases 1 to 800)  
 Bonaldo, M.F., Lennon, G. and Soares, M.B.  
 Normalization and subtraction: two approaches to facilitate gene  
 discovery  
 Genome Res. 6 (9), 791-806 (1996)  
 JOURNAL  
 MEDLINE  
 PUBMED  
 COMMENT

Contact: Soares, MB

Coordinated Laboratory for Computational Genomics  
 University of Iowa  
 375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA  
 Tel: 319 335 8250  
 Fax: 319 335 9565  
 Email: bento-soares@uiowa.edu  
 Tissue Procurement: Mary Hendrix  
 CDNA Library preparation: Dr. M. Bento Soares, University of Iowa  
 CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
 Clone Distribution: Distribution information can be found at  
<http://genome.uiowa.edu/distribution/humanfl.html>  
 The following repetitive elements were found in this CDNA  
 sequence: 172-253, >(G)A#Simple\_repeat (matched complement)  
 Seq primer: pYX-5,

## FEATURES

Location/Qualifiers  
 1..800  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:30554533"  
 /tissue\_type="Chondrosarcoma Lung Metastasis cell lines"  
 /lab\_host="DH10B (T1 phage resistant)"  
 /clone\_lib="NIH MGC 214"  
 /note="Organ: Lung; Vector: pYX-Asc; Site 1: EcoR I;  
 Site 2: Not I; The library was constructed according  
 Bonaldo, Lennon and Soares, Genome Research, 6:791-806,  
 1996. Denatured RNA was size fractionated on a 1% agarose

Genoscope - Centre National de Sequencage  
 BP 191 91006 Evry cedex - France  
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr  
 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime  
 end enriched, double-strand cDNA was digested with Not I and cloned  
 into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library  
 was normalized. Library was constructed by Life Technologies, a  
 division of Invitrogen. This sequence belongs to sequence cluster  
 3812.f

For more information about this cluster, see

http://www.genoscope.cns.fr/cdna?s=CSODI059CD090P1sc=3812.f.

## FEATURES

source  
 Location/Qualifiers  
 1..1099  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="CSODI059YH17"  
 /tissue\_type="PLACENTA COT 25-NORMALIZED"  
 /clone\_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"  
 /note="1st strand cDNA was primed with a NotI-oligo(dT)  
 primer. Five prime end enriched, double-strand cDNA was  
 digested with Not I and EcoR V sites of the Not I and EcoR V  
 sites of the pCMVSPORT 6 vector. Library was normalized."

## ORIGIN

Alignment Scores:  
 Pred. No.: 5,03e-152 Length: 1099  
 Score: 162.00 Matches: 162  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 90.50% Indels: 0  
 DB: 1 Gaps: 0

US-09-701-674A-23 (1-179) x AL552119 (1-1099)

QY 1 MetSerThrGlySerLeuSerAspValGluAspLeuGlnGluMetLeuGluCys 20  
 Db 260 ATGTCCACCGCTCCCTCAGCATGTGGAGGACCTTCAAGAGGTGGAGATGTTGAATGT 319  
 QY 21 AspGlyLeuMetAspSerAsnLysGluPheValThrSerAsnGluSerThrGluGlu 40  
 Db 320 GACGGGTGAAATGGATTCGAACAGGAATTTGTACTTCCACGAGAGACCGAGGAG 379  
 QY 41 SerSerAsnCysGluAsnGlySerProGlnLysGlyArgGlyGlyLeuGlyLysArgArg 60  
 Db 380 AGCTCCAACTGCAGAAATGGTCTCCCAAGAGGCGCGCGCTGGCAAGAGGAGG 439  
 QY 61 LysAlaProThrLysLysSerProLeuSerGlyValSerGlnGluGlyLysGlnValGln 80  
 Db 440 AAGGCGCCCAACCAAGAGAGCCCTGTAGCGGGGTCCAGCCAGGAGGGAAGCAGGTCCAG 499  
 QY 81 ArgAsnAlaAlaAsnAlaArgGluArgAlaArgMetArgValLeuSerLysAlaPheSer 100  
 Db 500 CGCAACCGCCCAACCGCGAGAGCGGGCCCGATCGGAGTTCTGAGCAAGGCCCTTCTCC 559  
 QY 101 ArgLeuLysThrThrLeuProTrpValProProAspThrLysLeuSerLysLeuAspThr 120  
 Db 560 AGACTCAAGACCAACCTCGCTGGGTGCGCCCGCACCAAGCTCTCCAAAGTGGACACG 619  
 QY 121 LeuArgLeuAlaSerSerTyrlleAlaHisLeuArgGlnIleLeuAlaAsnAspLysTyr 140  
 Db 620 CTCAGGCTGGGCTCCAGCTACATCGCCCACTTGAAGCAGATCCTGGCTAAACACAAATAC 679  
 QY 141 GluAsnGlyTyrlleHisProValAsnLeuThrTrpProPheMetValAlaGlyLysPro 160  
 Db 680 GAGAACGGGTACATTACCGGTCACTGACGTGGCCCTTTATGGTGGCGGGAAACCC 739  
 QY 161 GluSer 162  
 Db 740 GAGAGT 745

RESULT 10  
 AL552380

LOCUS AL552380 965 bp mRNA linear EST 30-MAR-2004  
 DEFINITION AL552380 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens CDNA  
 clone CSODI070YX16 5-PRIME, mRNA sequence.

ACCESSION AL552380

VERSION AL552380.3 GI:45857168

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE  
 AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

TITLE 1 (bases 1 to 965)

JOURNAL Li, W.B., Gruber, C., Jessee, J. and Polayes, D.

COMMENT Full-length cDNA libraries and normalization

Unpublished (2001)

On Feb 15, 2001 this sequence version replaced gi:31274195.

Contact: Genoscope

Genoscope - Centre National de Sequencage

BP 191 91006 Evry cedex - France

Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr

1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime  
 end enriched, double-strand cDNA was digested with Not I and cloned  
 into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library  
 was normalized. Library was constructed by Life Technologies, a  
 division of Invitrogen. This sequence belongs to sequence cluster  
 3812.f

For more information about this cluster, see

http://www.genoscope.cns.fr/cdna?s=CSODI070BF08QPlkc=3812.f.

## FEATURES

source

Location/Qualifiers

1..965

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="CSODI070YX16"

/tissue\_type="PLACENTA COT 25-NORMALIZED"

/clone\_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"

/note="1st strand cDNA was primed with a NotI-oligo(dT)  
 primer. Five prime end enriched, double-strand cDNA was  
 digested with Not I and EcoR V sites of the Not I and EcoR V  
 sites of the pCMVSPORT 6 vector. Library was normalized."

## ORIGIN

Alignment Scores:  
 Pred. No.: 4,56e-150 Length: 965  
 Score: 160.00 Matches: 175  
 Percent Similarity: 99.43% Conservative: 0  
 Best Local Similarity: 99.43% Mismatches: 1  
 Query Match: 89.39% Indels: 1  
 DB: 1 Gaps: 0

US-09-701-674A-23 (1-179) x AL552380 (1-965)

QY 1 MetSerThrGlySerLeuSerAspValGluAspLeuGlnGluValGluMetLeuGluCys 20

Db 272 ATGTCCACCGCTCCCTCAGCATGTGGAGGACCTTCAAGAGGTGGAGATGTTGAATGT 331

QY 21 AspGlyLeuLysMetAspSerAsnLysGluPheValThrSerAsnGluSerThrGluGlu 40

Db 332 GACGGGTGAAATGGATTCGAACAGGAATTTGTGACTTCCACGAGACCGAGGAG 391

QY 41 SerSerAsnCysGluAsnGlySerProGlnLysGlyArgGlyGlyLeuGlyLysArgArg 60

Db 392 AGCTCCAACTGCAGAAATGGTCTCTCCCAAGAGCGCGCGCTGGCAAGAGGAG 451

QY 61 LysAlaProThrLysLysSerProLeuSerGlyValSerGlnGluGlyLysGlnValGln 80

Db 452 AAGCGGCCCAACCAAGAGAGCCCTCAGCGGGGTACGCCAGAGGGGAAGCAGGTCCAG 511

QY 81 ArgAsnAlaAlaAsnAlaArgGluArgAlaArgMetArgValLeuSerLysAlaPheSer 100

Db 512 CGCAACCGCCCAACCGCGAGAGCGGCGCCCGATCGAGTGTGTGAGCAAGAGGCTTCTCC 571

QY 101 ArgLeuLysThrThrLeuProTrpValProProAspThrLysLeuSerLysLeuAspThr 120



```

/lab_host="DH10B (T1 phage resistant)"
/clone_lib="NIH_MGC_214"
/notes="Organ: Lung; Vector: pYX-Asc; Site 1: EcoR I;
Site 2: Not I; The library was constructed according
Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
1996. Denatured RNA was size fractionated on a 1% agarose
gel. First strand cDNA synthesis was primed with oligo-dT
primer containing a Not I site. Double strand cDNA was
size selected according to mRNA size fraction, ligated
with EcoR I adaptor, digested with Not I and then cloned
directionally into pYX-Asc vector. The library tag
sequence located between the Not I site and the polyA tail
is GATAGGCCCA. Tissue was provided by Mary Hendrix."

ORIGIN
Alignment Scores:
Pred. No.: 4,06e-141 Length: 757
Score: 151.00 Matches: 151
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 84.36% Indels: 0
DB: 6 Gaps: 0

US-09-701-674A-23 (1-179) x CF126927 (1-757)
QY 1 MetSerThrGlySerLeuSerAspValGluAspLeuGlnGluValGluMetLeuGluCys 20
Db 263 ATGTCCACCGGCTCCCTCAGCATGTGGAGGACCTTCAAGAGGTGGAGATGTGGAATGT 322
QY 21 AspGlyLeuMetAspSerLeuSerGluPheValThrSerAsnGluSerThrGluGlu 40
Db 323 GACGGGTGAAAATGGATTGCAACAGGAATTTGTGACTTCCACAGAGACCCGAGGAG 382
QY 41 SerSerAsnGlySerProGlnLysGlyArgGlyGlyLeuGlyLysArgArg 60
Db 383 AGCTCAACTGCGAGAAATGGTCTCCCAAGAGGCGCGCGCTGGGCAAGAGGAGG 442
QY 61 LysAlaProThrLysSerProLeuSerGlyValSerGlnGluGlyLysGlnValGln 80
Db 443 AAGCGCCCAACCAAGAGAGCCCTCGAGCGGGTCCAGCCAGGAGGAGGAGTCCAG 502
QY 81 ArgAsnAlaAlaAsnAlaArgGluArgAlaArgMetArgValLeuSerLysAlaPheSer 100
Db 503 CGCAACGCGCCACACGCGGAGAGCGGGCCGCAATGCGAGTCTCGACCAAGCCCTTCTCC 562
QY 101 ArgLeuLysThrThrLeuProThrValProProAspThrLysLeuSerLysLeuAspThr 120
Db 563 AGACTCAAGACCAACCTCCCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 622
QY 121 LeuArgLeuAlaSerSerTyrrileAlaHisLeuArgGlnIleLeuAlaAsnAspLysTyrr 140
Db 623 CTCAGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 682
QY 141 GluAsnGlyTyrrileHisProValAsnLeuLeuThr 151
Db 683 GAGAACGGGTACATTCAACCGGGTCAACCTGACG 715

RESULT 13
BI770998 746 bp mRNA linear EST 25-SEP-2001
LOCUS 60305534:1 NIH_MGC_122 Homo sapiens cDNA clone IMAGE:5204651 5',
DEFINITION mRNA sequence.
ACCESSION BI770998
VERSION BI770998
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 746)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)

Contact: Robert Strausberg, Ph.D.
Email: ccapbs-remail.nih.gov
Tissue Procurement: Life Technologies, Inc.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1AM1513 row: b column: 12
High quality sequence stop: 744.
Location/Qualifiers
1..746
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5204651"
/lab_host="DH10B"
/clone_lib="NIH_MGC_122"
/notes="Organ: pooled lung and spleen; Vector: pCMV-SPORT6;
Site 1: NotI; Site 2: EcoRV (destroyed); RNA source
anonymous pool of 24 week female lung, 16 week female
spleen, and 20-22 week male spleens. Library is oligo-dT
primed and directionally cloned (EcoRV site is destroyed
upon cloning). Average insert size 1.4 kb, insert size
range 1-3 kb. Library is normalized and enriched for
full-length clones and was constructed by C. Gruber
(Invitrogen). Research Genetics tracking code 026. Note:
this is a NIH_MGC Library."

ORIGIN
Alignment Scores:
Pred. No.: 4,1e-139 Length: 746
Score: 149.00 Matches: 162
Percent Similarity: 99.39% Conservative: 0
Best Local Similarity: 99.39% Mismatches: 0
Query Match: 83.24% Indels: 1
DB: 4 Gaps: 0

US-09-701-674A-23 (1-179) x BI770998 (1-746)
QY 1 MetSerThrGlySerLeuSerAspValGluAspLeuGlnGluValGluMetLeuGluCys 20
Db 258 ATGTCCACCGGCTCCCTCAGCATGTGGAGGACCTTCAAGAGGTGGAGATGTGGAATGT 317
QY 21 AspGlyLeuLysMetAspSerAsnLysGluPheValThrSerAsnGluSerThrGluGlu 40
Db 318 GACGGGTGAAAATGGATTGCAACAGGAATTTGTGACTTCCACAGAGACCCGAGGAG 377
QY 41 SerSerAsnGlySerProGlnLysGlyArgGlyGlyLeuGlyLysArgArg 60
Db 378 AGCTCAACTGCGAGAAATGGTCTCCCAAGAGGCGCGCGCTGGGCAAGAGGAGG 437
QY 61 LysAlaProThrLysLysSerProLeuSerGlyValSerGlnGluGlyLysGlnValGln 80
Db 438 AAGCGCCCAACCAAGAGAGCCCTCGAGCGGGTCCAGCCAGGAGGAGGAGGAGTCCAG 497
QY 81 ArgAsnAlaAlaAsnAlaArgGluArgAlaArgMetArgValLeuSerLysAlaPheSer 100
Db 498 CGCAACCGCCCAACGCGGAGAGCGGGCCGCAATGCGAGTCTCGACCAAGCCCTTCTCC 557
QY 101 ArgLeuLysThrThrLeuProThrValProProAspThrLysLeuSerLysLeuAspThr 120
Db 558 AGACTCAAGACCAACCTCCCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 617
QY 121 LeuArgLeuAlaSerSerTyrrileAlaHisLeuArgGlnIleLeuAlaAsnAspLysTyrr 140
Db 618 CTCAGGTGGGTCAGCTACATCGCCCACTTGAAGGAGATCTCTGGTGAACGACAAATAC 677
QY 141 GluAsnGlyTyrrileHisProValAsnLeuLeuThrProPheMetValAlaGlyLysPr 160
Db 678 GAGAACGGGTACATTCAACCGGGTCAACCTGACGCTTATGTGGTGGCGGGAACCC 737

```

```

QY      160 oGluSer 162
Db      738 CGAGAGT 744

RESULT 14
CF128246
LOCUS   645 bp mRNA linear EST 05-AUG-2003
DEFINITION
UI-HF-ET0-awf-c-24-0-UI.r1 NIH_MGC_214 Homo sapiens cDNA clone
IMAGE:30555623 5', mRNA sequence.
ACCESSION
CF128246
VERSION
CF128246.1 GI:33207297
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 645)
AUTHORS
Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE
Normalization and subtraction: two approaches to facilitate gene
discovery
JOURNAL
Genome Res. 6 (9), 791-806 (1996)
MEDLINE
97044477
PUBMED
8889548
COMMENT
Contact: Soares, MB
Coordinated Laboratory for Computational Genomics
University of Iowa
375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: bento-soares@uiowa.edu
Tissue Procurement: Mary Hendrix
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Distribution information can be found at
http://genome.uiowa.edu/distribution/humanfl.html
The following repetitive elements were found in this cDNA
sequence: 39-120, >(GA)n#Simple_repeat (matched complement)
Seq primer: pYX-5

FEATURES
source
Location/Qualifiers
1..645
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:30555623"
/tissue_type="Chondrosarcoma Lung Metastasis cell lines"
/lab_host="DH10B (T1 phage resistant)"
/clone_lib="NIH_MGC_214"
/notes="Organ: Lung; Vector: pYX-Asc; Site_1: EcoR I;
Site_2: Not I; The library was constructed according
Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
1996. Denatured RNA was size fractionated on a 1% agarose
gel. First strand cDNA synthesis was primed with oligo-dT
primer containing a Not I site. Double strand cDNA was
size selected according to mRNA size fraction, ligated
with EcoR I adaptor, digested with Not I and then cloned
directionally into pYX-Asc vector. The library tag
sequence located between the Not I site and the polyA tail
is GATAGGCCA. Tissue was provided by Mary Hendrix."

ORIGIN
Alignment Scores:
Pred. No.: 4,08e-128 Length: 645
Score: 138.00 Matches: 138
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 77.09% Indels: 0
DB: 6 Gaps: 0

US-09-701-674A-23 (1-179) x CF128246 (1-645)

QY      1 MetSerThrGlySerLeuSerAspValGluAspLeuGlnGluValGluMetLeuGluCys 20
|||||

```

```

Db      125 ATGTCCACCGCTCCCTCAGCGATGTGGAGGACCTTCAAGAGGTGGAGATGTGGATGT 184
QY      21 AspGlyLeuLysMetAspSerAsnLysGluPheValThrSerAsnGluSerThrGluGlu 40
|||||
Db      185 GACGGGTTGAAATGGATTCCAAAGGAATTGTGACTTCCAAACGAGACCCGAGGAG 244
QY      41 SerSerAsnCysGluAsnGlySerProGlnLysGlyArgGlyGlyGlyLysArgGAG 60
|||||
Db      245 AGCTCCAACTGCCAGATGGGTCTCCCAAGAGGGCGCGCGCTGGCAAGAGAGG 304
QY      61 LysAlaProThrLysLysSerProLeuSerGlyValSerGlnGluGlyLysGlnValGln 80
|||||
Db      305 AAGCGGCCCAACCAAGAGAGCCCTCAGCGGGGTCCAGCCAGGAGGGAGAGGTCCAG 364
QY      81 ArgAsnAlaAlaAsnAlaArgGluArgAlaArgMetArgValLeuSerLysAlaPheSer 100
|||||
Db      365 CGCAACGCCGCCCAACCGCGAGAGCGCGCGCATCGAGTGTCTGAGCAAGGCGCTTCTCC 424
QY      101 ArgLeuLysThrThrLeuProTTPValProProAspThrLysLeuSerLysLeuAspThr 120
|||||
Db      425 AGACTCAAGACCACTCCCTGGTGGTCCCGCCGACCAAGCTCTCCAAAGCTGGACAG 484
QY      121 LeuArgLeuAlaSerSerTyrIleAlaHisLeuArgGlnIleLeuAlaAsnAsp 138
|||||
Db      485 CTCAGGCTGGCGTCCAGCTACATCGCCCACTTGAGGCAGATCTCGCTAACGAC 538

RESULT 15
CF128914
LOCUS   744 bp mRNA linear EST 05-AUG-2003
DEFINITION
UI-HF-ETO-avu-p-09-0-UI.r1 NIH_MGC_214 Homo sapiens cDNA clone
IMAGE:30563216 5', mRNA sequence.
ACCESSION
CF128914
VERSION
CF128914.1 GI:33208641
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 744)
AUTHORS
Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE
Normalization and subtraction: two approaches to facilitate gene
discovery
JOURNAL
Genome Res. 6 (9), 791-806 (1996)
MEDLINE
97044477
PUBMED
8889548
COMMENT
Contact: Soares, MB
Coordinated Laboratory for Computational Genomics
University of Iowa
375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: bento-soares@uiowa.edu
Tissue Procurement: Mary Hendrix
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Distribution information can be found at
http://genome.uiowa.edu/distribution/humanfl.html
The following repetitive elements were found in this cDNA
sequence: 191-272, >(GA)n#Simple_repeat (matched complement)
Seq primer: pYX-5

FEATURES
source
Location/Qualifiers
1..744
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:30563216"
/tissue_type="Chondrosarcoma Lung Metastasis cell lines"
/lab_host="DH10B (T1 phage resistant)"
/clone_lib="NIH_MGC_214"
/notes="Organ: Lung; Vector: pYX-Asc; Site_1: EcoR I;
Site_2: Not I; The library was constructed according
Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
1996. Denatured RNA was size fractionated on a 1% agarose
gel. First strand cDNA synthesis was primed with oligo-dT
primer containing a Not I site. Double strand cDNA was
size selected according to mRNA size fraction, ligated
with EcoR I adaptor, digested with Not I and then cloned
directionally into pYX-Asc vector. The library tag
sequence located between the Not I site and the polyA tail
is GATAGGCCA. Tissue was provided by Mary Hendrix."

ORIGIN
Alignment Scores:
Pred. No.: 4,08e-128 Length: 645
Score: 138.00 Matches: 138
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 77.09% Indels: 0
DB: 6 Gaps: 0

US-09-701-674A-23 (1-179) x CF128246 (1-645)

QY      1 MetSerThrGlySerLeuSerAspValGluAspLeuGlnGluValGluMetLeuGluCys 20
|||||

```

1996. Denatured RNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with oligo-dT primer containing a Not I site. Double strand cDNA was size selected according to mRNA size fraction, ligated with EcoR I adaptor, digested with Not I and then cloned directionally into pX-Asc vector. The library tag sequence located between the Not I site and the polyA tail is GATAAGGCCA. Tissue was provided by Mary Hendrix."

## ORIGIN

## Alignment Scores:

Pred. No.:	4.65e-128	Length:	744
Score:	138.00	Matches:	138
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	77.09%	Indels:	0
DB:	6	Gaps:	0

US-09-701-674A-23 (1-179) x CF128914 (1-744)

QY	1	MetSerThrGlySerLeuSerAspValGluAspLeuGlnGluValGluMetLeuGluCys	20
DB	277	ATGTCACCGCGTCCCTCAGCATGTGGAGGACCTTCAAGAGGTGGAGATGTGGAATGT	336
QY	21	AspGlyLeuLysMetAspSerAsnLysGluPheValThrSerAsnGluSerThrGluGlu	40
DB	337	GACGGGTTGAAATGATTGGAACAGAGGAATTTGTGACTTCCACGAGAGCACCAGGAG	396
QY	41	SerSerAsnCysGluAsnGlySerProGlnLysGlyArgGlyLeuGlyLysArgArg	60
DB	397	AGCTCCAACTGGAGAAATGGGTCTCCCCAGAAAGGGCGCGCGCTTGGGCAAGAGAGG	456
QY	61	LysAlaProThrLysLysSerProLeuSerGlyValSerGlnGluGlyLysGlnValGln	80
DB	457	AAGGCGCCCAACAGAGAGGCCCCCTGAGCGGGTCCAGCAGAGGGGAGCGGTCCAG	516
QY	81	ArgAsnAlaAlaAsnAlaArgGluArgAlaArgMetArgValLeuSerLysAlaPheSer	100
DB	517	CGCAGCGCGCCCAACCGCGAGAGCGGGCCGCGATGCGAGTGTGAGCAAGGCCCTCTCC	576
QY	101	ArgLeuLysThrThrLeuProTrpValProProAspThrLysLeuSerLysLeuAspThr	120
DB	577	AGACTCAGACCAACCTCGCTGGTGGTCCCCCGACACCAAGCTCTCCAAAGCTGGACAG	636
QY	121	LeuArgLeuAlaSerSerTyrIleAlaHisLeuArgGlnIleLeuAlaAsnAsp	138
DB	637	CTCAGGCTGGGTCCAGCTACATCGCCCACTTGAGGCAGATCCTGGCTAACGAC	690

Search completed: December 17, 2004, 03:13:22  
Job time : 2874 secs

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: December 16, 2004, 23:16:02 ; Search time 487 Seconds  
(without alignments)  
2029.220 Million cell updates/sec

Title: US-09-701-674A-23

Perfect score: 917

Sequence: 1 MGTGSLSDVEDIQEVMLEEC.....PESDLKEVTVASRLGGTTAS 179

Scoring table:

BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 4093002 seqs, 2760418825 residues

Total number of hits satisfying chosen parameters: 8186004

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlh  
-O=/cgn2\_1/USPTO.spool\_h/US09701674/runat\_15122004\_100443\_27616/app.query.fasta\_1.327  
-DB=Published Applications NA -QWTF=fastap -SUFFIX=rnpb -MINMATCH=0.1  
-LOOPL=0 -LOOPEXT=0 -UNIT8=bits -START=1 -THR=SCORE=pct -BLOSUM62  
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -END=1 -MATRIX=THRMAX=100  
-THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pct -NORM=ext -HEAPSIZ=500 -MINLEN=0  
-MAXLEN=2000000000 -USER=US09701674@cgn\_1\_1480@runat\_15122004\_100443\_27616  
-NCPUS=6 -ICPU=3 -NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOC=100  
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5  
-FGAPOP=6 -FGAPEXT=7 -XGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA:

1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq:  
2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq:  
3: /cgn2\_6/ptodata/1/pubpna/US05\_NEW\_PUB.seq:  
4: /cgn2\_6/ptodata/1/pubpna/US05\_PUBCOMB.seq:  
5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq:  
6: /cgn2\_6/ptodata/1/pubpna/PCTUS\_PUBCOMB.seq:  
7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq:  
8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq:  
9: /cgn2\_6/ptodata/1/pubpna/US09A\_PUBCOMB.seq:  
10: /cgn2\_6/ptodata/1/pubpna/US09B\_PUBCOMB.seq:  
11: /cgn2\_6/ptodata/1/pubpna/US09C\_PUBCOMB.seq:  
12: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq:  
13: /cgn2\_6/ptodata/1/pubpna/US10A\_PUBCOMB.seq:  
14: /cgn2\_6/ptodata/1/pubpna/US10B\_PUBCOMB.seq:  
15: /cgn2\_6/ptodata/1/pubpna/US10C\_PUBCOMB.seq:  
16: /cgn2\_6/ptodata/1/pubpna/US10D\_PUBCOMB.seq:  
17: /cgn2\_6/ptodata/1/pubpna/US10E\_PUBCOMB.seq:  
18: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq:  
19: /cgn2\_6/ptodata/1/pubpna/US11\_NEW\_PUB.seq:  
20: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq:  
21: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

TYPE: DNA

ORGANISM: Homo sapiens

FEATURE:

Result No.	Score	Match	Length	DB ID	Description
1	764.5	83.4	632	16	US-10-264-049-1307
2	514	56.1	2177	15	US-10-247-671-97
3	514	56.1	2196	15	US-10-084-817-120
4	507	55.3	2382	15	US-10-240-965-83
5	493.5	53.8	1716	17	US-10-775-169-201
6	487	53.1	1939	17	US-09-918-995-9069
7	482	52.6	1939	17	US-10-755-889-365
8	353	38.5	240	16	US-10-305-720-1028
9	318	34.7	697	9	US-09-954-456-829
10	318	34.7	697	9	US-09-954-456-1245
11	318	34.7	697	10	US-09-960-706-1028
12	318	34.7	697	10	US-09-873-313-679
13	228	24.9	493	10	US-09-918-995-11181
14	180.5	19.7	1140	17	US-10-450-826-124
15	168.5	18.4	446	9	US-09-833-381-505
16	165.5	18.0	651	9	US-09-749-728B-22
17	165.5	18.0	2465	13	US-10-044-090-251
18	164	17.9	609	15	US-10-029-386-22740
19	164	17.9	1800	10	US-09-771-357-106
20	164	17.9	1800	15	US-10-059-579-106
21	162	17.7	1457	9	US-09-954-531-982
22	162	17.7	1467	16	US-10-062-674-1950
23	162	17.7	1678	15	US-10-240-965-119
24	157.5	17.2	907	13	US-10-004-717-65
25	157.5	17.2	907	18	US-10-860-373-65
26	157.5	17.2	907	18	US-10-860-724-65
27	157.5	17.2	2537	16	US-10-466-164-25
28	157.5	17.2	2550	18	US-10-335-053-106
29	157.5	17.2	2601	9	US-09-880-107-2433
30	157	17.1	1477	13	US-10-004-717-63
31	157	17.1	1477	18	US-10-860-373-63
32	157	17.1	1477	18	US-10-860-724-63
33	151.5	16.5	501	13	US-10-004-717-13
34	151.5	16.5	501	18	US-10-860-373-13
35	151.5	16.5	501	18	US-10-860-724-13
36	151	16.5	515	13	US-10-004-717-35
37	151	16.5	515	18	US-10-860-373-35
38	151	16.5	515	18	US-10-860-724-35
39	150.5	16.4	4153	13	US-10-116-275-194
40	148.5	16.2	993	13	US-10-004-717-47
41	148.5	16.2	993	18	US-10-860-373-47
42	148.5	16.2	993	18	US-10-860-724-47
43	148.5	16.2	3261	13	US-10-004-717-12
44	148.5	16.2	3261	18	US-10-860-373-12
45	148.5	16.2	3261	18	US-10-860-724-12

#### ALIGNMENTS

RESULT 1  
US-10-264-049-1307  
; Sequence 1307, Application US/10264049  
; Publication No. US20040005579A1  
; GENERAL INFORMATION:  
; APPLICANT: Birse et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PA133P1  
; CURRENT APPLICATION NUMBER: US/10/264,049  
; CURRENT FILING DATE: 2002-10-04  
; PRIOR APPLICATION NUMBER: PCT/US01/18569  
; PRIOR FILING DATE: 2001-06-07  
; PRIOR APPLICATION NUMBER: US 60/209,467  
; PRIOR FILING DATE: 2000-06-07  
; NUMBER OF SEQ ID NOS: 4360  
; SOFTWARE: PatentIn Ver. 3.1  
; SEQ ID NO 1307  
; LENGTH: 632  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:



```

1 NAME/KEY: misc_feature
2 LOCATION: (78)..(79)
3 OTHER INFORMATION: n equals a,t,g, or c
4 FEATURE:
5 NAME/KEY: misc_feature
6 LOCATION: (174)..(174)
7 OTHER INFORMATION: n equals a,t,g, or c
8 FEATURE:
9 NAME/KEY: misc_feature
10 LOCATION: (231)..(231)
11 OTHER INFORMATION: n equals a,t,g, or c
12 FEATURE:
13 NAME/KEY: misc_feature
14 LOCATION: (249)..(249)
15 OTHER INFORMATION: n equals a,t,g, or c
16 FEATURE:
17 NAME/KEY: misc_feature
18 LOCATION: (263)..(263)
19 OTHER INFORMATION: n equals a,t,g, or c
20 FEATURE:
21 NAME/KEY: misc_feature
22 LOCATION: (265)..(265)
23 OTHER INFORMATION: n equals a,t,g, or c
24 FEATURE:
25 NAME/KEY: misc_feature
26 LOCATION: (280)..(280)
27 OTHER INFORMATION: n equals a,t,g, or c
28 FEATURE:
29 NAME/KEY: misc_feature
30 LOCATION: (331)..(331)
31 OTHER INFORMATION: n equals a,t,g, or c
32 FEATURE:
33 NAME/KEY: misc_feature
34 LOCATION: (433)..(433)
35 OTHER INFORMATION: n equals a,t,g, or c
36 FEATURE:
37 NAME/KEY: misc_feature
38 LOCATION: (467)..(467)
39 OTHER INFORMATION: n equals a,t,g, or c
40 FEATURE:
41 NAME/KEY: misc_feature
42 LOCATION: (498)..(498)
43 OTHER INFORMATION: n equals a,t,g, or c
44 FEATURE:
45 NAME/KEY: misc_feature
46 LOCATION: (537)..(537)
47 OTHER INFORMATION: n equals a,t,g, or c
48 FEATURE:
49 NAME/KEY: misc_feature
50 LOCATION: (546)..(546)
51 OTHER INFORMATION: n equals a,t,g, or c
52 FEATURE:
53 NAME/KEY: misc_feature
54 LOCATION: (552)..(552)
55 OTHER INFORMATION: n equals a,t,g, or c
56 FEATURE:
57 NAME/KEY: misc_feature
58 LOCATION: (623)..(623)
59 OTHER INFORMATION: n equals a,t,g, or c
60 FEATURE:
61 NAME/KEY: misc_feature
62 LOCATION: (629)..(629)
63 OTHER INFORMATION: n equals a,t,g, or c
64 PS-10-264-049-1307

```

US-09-701-674A-23 (1-179) x US-10-264-049-1307 (1-632)

QY 1 MetSerThrGlySerLeuSerAspValGluAspLeuGlnGluValGluMetLeuGlnCys 20  
Db 13 ATGTCACCGGCTCCCTTCAGCGATGTGGAGGACCTTCAAGAGGTGGAGATGTGTGAATGT 72  
QY 21 AspGlyLeuLeuMetAspSerAsnLysGluPheValThrSerAsnGluSerThrGluGlu 40  
Db 73 GACGNNNTGAATATGGATTTCGACAGGAATTTGTGACCTTCCACGAGAGCACCGAGGAG 132  
QY 41 SerSerAsnCysGluAsnGlySerProGlnLysGlyArgGlyClyLeuGlyLysArgArg 60  
Db 133 AGCTCCAACTCGGAGATGGTCTCTCCACAGAGGGCCGGGCGCTGGGCGCAAGAGGAGG 192  
QY 61 LysAlaProThrLysLysSerProLeuSerGlyValSerGlnClyGlyLysGlnValGln 80  
Db 193 AAGCGCGCCCAACAAGAGAGCCCCCTGAGCGGGTGTCAAGCAGAGGGGAGCAGGTCAG 252  
QY 81 ArgAsnAlaAlaAsnAlaArgGluArgAlaArgMetArgValLeuSerLysAlaPheSer 100  
Db 253 CGCAACGCGCGNACGCGCGAGAGCGGCGCATCGAGTGTGCTAGCAGCAAGGCCCTTC 312  
QY 101 ArgLeuLysThrThrLeuProTTPValProProAspThrLysLeuSerLysLeuAspThr 120  
Db 313 AGACTCAAGACACCCCTGNCCTGGTGGTCCCCCGACCAACCAAGCTCTCCAAGCTGGAC 372  
QY 121 LeuArgLeuAlaSerSerTyrIleAlaHisLeuArgGlnIleLeuAlaAsnAspLysTyr 140  
Db 373 CTCAGCTGGCGTCCAGCTACATCGCCCACTTCGAGCGAGATCCTGGCTTAACGACCA 432  
QY 141 GluAsnGlyTyrIleHisProValAsnLeuThrTTPProPheMetValAlaGlyLysPro 160  
Db 433 NAGAACGGGTACATTA--CCGGGCAACTGACGTGGNCCCTTTATGGTGGC-GGGAAACC 489  
QY 161 GluSerAspLeuLysGluValVal-ThrAlaSerArgLeuLysGlyThrThrAla 178  
Db 490 GAGAGTGANTGAAAGAGTGGTTGACCGCGAGCGCGGTTATGTGGAANCACCGCG 544

RESULT 2  
US-10-247-671-97  
; Sequence 97, Application US/10247671  
; Publication No. US20030194721A1  
; GENERAL INFORMATION:  
; APPLICANT: Mikita, Thomas  
; APPLICANT: Shiffman, Dov  
; APPLICANT: Porter, Gordon, J.  
; APPLICANT: Kaser, Matthew R.  
; TITLE OF INVENTION: GENES EXPRESSED IN TREATED FOAM CELLS  
; FILE REFERENCE: PA-0050 US  
; CURRENT APPLICATION NUMBER: US/10/247,671  
; CURRENT FILING DATE: 2002-09-18  
; PRIOR APPLICATION NUMBER: 60/323,784  
; PRIOR FILING DATE: 2001-09-19  
; NUMBER OF SEQ ID NOS: 186  
; SOFTWARE: PERL Program  
; SEQ ID NO 97  
; LENGTH: 2177  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; OTHER INFORMATION: Incyte ID NO. US20030194721A1 3282941.CB1  
; US-10-247-671-97

US-09-701-674A-23 (1-179) x US-10-247-671-97 (1-2177)







Score: 353.00 Matches: 76  
 Percent Similarity: 93.83% Conservative: 0  
 Best Local Similarity: 93.83% Mismatches: 3  
 Query Match: 38.50% Indels: 2  
 DB: 16 Gaps: 0

US-09-701-674A-23 (1-179) x US-10-305-720-1028 (1-240)

Qy 61 LysAlaProThrLysLysSerProLeuSerGlyValSerGln-GluGlyLysGlnValG1 80  
 Db 1 AAGGCGCCCAAGAGAGCCCTGAGCGGGGTGAGCGAGGAGGGGAGCAGGTCCA 60  
 Qy 80 nAtgAsnAlaAlaAsnAlaArgGluArg-AlaArgMetArgValLeuSerLysAlaPheS 100  
 Db 61 GCGCAAGCNGCCCAAGCGCGAGAGCGGCGCCGCAATGCGAGTGTGAGCAAGGCTTCT 120  
 Qy 100 erArgLeuLysThrThrLeuProToTrpValProProAspThrLysLeuSerLysLeuAspT 120  
 Db 121 CCAGACTCAAGACCAACCTGCGCTGGTGGTCCCGCCGACACCAAGCTCTCCAAGCTGGACA 180  
 Qy 120 hrLeuArgLeuAlaSerSerTyrIleAlaHisLeuArgGlnIleLeuAlaHisLeuAspLys 139  
 Db 181 CGCTCAGGCTGGGCTCCAGTACATNGCCCACTTGAGGCGAGATCTCTGGCTACAGCAAA 239

# RESULT 9

US-09-954-456-829  
 ; Sequence 829, Application US/09954456  
 ; Patent No. US20020115057A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Young, Paul  
 ; TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using Canc

FILE REFERENCE: 689290-76  
 CURRENT APPLICATION NUMBER: US/09/954,456  
 CURRENT FILING DATE: 2001-09-18  
 PRIOR APPLICATION NUMBER: US/60/233,617  
 PRIOR FILING DATE: 2000-09-18  
 PRIOR APPLICATION NUMBER: US/60/234,052  
 PRIOR FILING DATE: 2000-09-20  
 PRIOR APPLICATION NUMBER: US/60/234,923  
 PRIOR FILING DATE: 2000-09-25  
 PRIOR APPLICATION NUMBER: US/60/235,134  
 PRIOR FILING DATE: 2000-09-25  
 PRIOR APPLICATION NUMBER: US/60/235,637  
 PRIOR FILING DATE: 2000-09-26  
 PRIOR APPLICATION NUMBER: US/60/235,638  
 PRIOR FILING DATE: 2000-09-26  
 PRIOR APPLICATION NUMBER: US/60/235,711  
 PRIOR FILING DATE: 2000-09-27  
 PRIOR APPLICATION NUMBER: US/60/235,720  
 PRIOR FILING DATE: 2000-09-27  
 PRIOR APPLICATION NUMBER: US/60/235,840  
 PRIOR FILING DATE: 2000-09-27  
 PRIOR APPLICATION NUMBER: US/60/235,863  
 PRIOR FILING DATE: 2000-09-27  
 NUMBER OF SEQ ID NOS: 2276  
 SOFTWARE: PatentIn version 3.0  
 SEQ ID NO 829  
 LENGTH: 697  
 TYPE: DNA  
 ORGANISM: Homo sapiens  
 FEATURE:  
 NAME/KEY: misc feature  
 OTHER INFORMATION: n=a,t,g or c  
 US-09-954-456-829

Alignment Scores:  
 Pred. No.: 3,546-30 Length: 697  
 Score: 318.00 Matches: 61  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 34.68% Indels: 0  
 DB: 9 Gaps: 0

US-09-701-674A-23 (1-179) x US-09-954-456-829 (1-697)

Qy 119 AspThrLeuArgLeuAlaSerSerTyrIleAlaHisLeuArgGlnIleLeuAlaAsnApp 138  
 Db 3 GACACGCTCAGGCTGGGCTCAGCTACATCGCCCACTTGAGGCGAGATCTCTGGCTAACGAC 62  
 Qy 139 LysTyrGluAsnGlyTyrIleHisProValAsnLeuThrTrpProPheValAlaGly 158  
 Db 63 AATACAGAGAACGGGTACATTCACCGGTCAACTGACGCGCCCTTATGGTGGCGGG 122  
 Qy 159 LysProGluSerAspLeuLysGluValValThrAlaSerArgLeuCysGlyThrThrAla 178  
 Db 123 AAACCCGAGAGTGACCTGAAAGAAGTGGTGACCGCGAGCGCTTATGTGGACCAACCGCG 182  
 Qy 179 Ser 179  
 Db 183 TCC 185

# RESULT 10

US-09-954-456-1245  
 ; Sequence 1245, Application US/09954456  
 ; Patent No. US20020115057A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Young, Paul  
 ; TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using Canc

FILE REFERENCE: 689290-76  
 CURRENT APPLICATION NUMBER: US/09/954,456  
 CURRENT FILING DATE: 2001-09-18  
 PRIOR APPLICATION NUMBER: US/60/233,617  
 PRIOR FILING DATE: 2000-09-18  
 PRIOR APPLICATION NUMBER: US/60/234,052  
 PRIOR FILING DATE: 2000-09-20  
 PRIOR APPLICATION NUMBER: US/60/234,923  
 PRIOR FILING DATE: 2000-09-25  
 PRIOR APPLICATION NUMBER: US/60/235,134  
 PRIOR FILING DATE: 2000-09-25  
 PRIOR APPLICATION NUMBER: US/60/235,637  
 PRIOR FILING DATE: 2000-09-26  
 PRIOR APPLICATION NUMBER: US/60/235,638  
 PRIOR FILING DATE: 2000-09-26  
 PRIOR APPLICATION NUMBER: US/60/235,711  
 PRIOR FILING DATE: 2000-09-27  
 PRIOR APPLICATION NUMBER: US/60/235,720  
 PRIOR FILING DATE: 2000-09-27  
 PRIOR APPLICATION NUMBER: US/60/235,840  
 PRIOR FILING DATE: 2000-09-27  
 PRIOR APPLICATION NUMBER: US/60/235,863  
 PRIOR FILING DATE: 2000-09-27  
 NUMBER OF SEQ ID NOS: 2276  
 SOFTWARE: PatentIn version 3.0  
 SEQ ID NO 1245  
 LENGTH: 697  
 TYPE: DNA  
 ORGANISM: Homo sapiens  
 FEATURE:  
 NAME/KEY: misc feature  
 OTHER INFORMATION: n=a,t,g or c  
 US-09-954-456-1245

Alignment Scores:  
 Pred. No.: 3,546-30 Length: 697  
 Score: 318.00 Matches: 61  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 34.68% Indels: 0  
 DB: 9 Gaps: 0

US-09-701-674A-23 (1-179) x US-09-954-456-1245 (1-697)

Qy 119 AspThrLeuArgLeuAlaSerSerTyrIleAlaHisLeuArgGlnIleLeuAlaAsnApp 138

Db 3 GACACGCTCAGGTCGGTCAGCTACATCGCCACTTGAGCGAGATCCTGGCTAACGAC 62  
 QY 139 LysTyrGluAsnGlyTyrIleHisProValAsnLeuThrTrpProPheMetValAlaGly 158  
 Db 63 AAATACGAGAACGGGTACATTACCCGGTCAACCTGAGCTGGCCCTTTATGTGGCCGGG 122  
 QY 159 LysProGluSerAspLeuLysGluValThrAlaSerArgLeuCysGlyThrThrAla 178  
 Db 123 AAACCCGAGAGTGACCTGAAAGAGTGGTGACCGGAGCGCGCTTATGTGGAACCCGCG 182  
 QY 179 Ser 179  
 Db 183 TCC 185

## RESULT 11

US-09-960-706-1028  
 ; Sequence 1028, Application US/09960706  
 ; Publication No. US20030134280A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Munger, William E.  
 ; TITLE OF INVENTION: Identifying Drugs for and Diagnosis of Benign Prostatic Hyperplasia  
 ; FILE REFERENCE: 44921-5029-01US  
 ; CURRENT APPLICATION NUMBER: US/09/960,706  
 ; CURRENT FILING DATE: 2001-09-24  
 ; PRIOR APPLICATION NUMBER: 60/223,323  
 ; PRIOR FILING DATE: 2000-08-07  
 ; PRIOR APPLICATION NUMBER: 09/873,319  
 ; PRIOR FILING DATE: 2001-06-05  
 ; NUMBER OF SEQ ID NOS: 1124  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 1028  
 ; LENGTH: 697  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; OTHER INFORMATION: Genbank Accession No. US20030134280A1 W73859  
 ; NAME/KEY: unsure  
 ; LOCATION: (1)..(697)  
 ; OTHER INFORMATION: n = a o r c o r g o r t  
 US-09-960-706-1028

Alignment Scores:  
 Pred. No.: 3,54e-30 Length: 697  
 Score: 318.00 Matches: 61  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 34.68% Indels: 0  
 DB: 10 Gaps: 0

US-09-701-674A-23 (1-179) x US-09-960-706-1028 (1-697)

QY 119 AspThrLeuArgLeuAlaSerSerTyrIleAlaHisLeuArgGlnIleLeuAlaAsnAsp 138  
 Db 3 GACACGCTCAGGTCGGTCAGCTACATCGCCACTTGAGCGAGATCCTGGCTAACGAC 62  
 QY 139 LysTyrGluAsnGlyTyrIleHisProValAsnLeuThrTrpProPheMetValAlaGly 158  
 Db 63 AAATACGAGAACGGGTACATTACCCGGTCAACCTGAGCTGGCCCTTTATGTGGCCGGG 122  
 QY 159 LysProGluSerAspLeuLysGluValThrAlaSerArgLeuCysGlyThrThrAla 178  
 Db 123 AAACCCGAGAGTGACCTGAAAGAGTGGTGACCGGAGCGCGCTTATGTGGAACCCGCG 182  
 QY 179 Ser 179  
 Db 183 TCC 185

## RESULT 12

US-09-873-319-679  
 ; Sequence 679, Application US/09873319A  
 ; Publication No. US20030134324A1  
 ; GENERAL INFORMATION:

; APPLICANT: Munger, William E.  
 ; APPLICANT: Kulkarni, Prakash  
 ; APPLICANT: Getzenberg, Robert H.  
 ; APPLICANT: Waga, Iwao  
 ; APPLICANT: Yamamoto, Jun  
 ; TITLE OF INVENTION: Identifying Drugs for and Diagnosis of Benign Prostatic  
 ; FILE REFERENCE: 44921-5029-05  
 ; CURRENT APPLICATION NUMBER: US/09/873,319A  
 ; CURRENT FILING DATE: 2001-06-05  
 ; EARLIER APPLICATION NUMBER: US 60/223,323  
 ; EARLIER FILING DATE: 2000-08-07  
 ; NUMBER OF SEQ ID NOS: 755  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 679  
 ; LENGTH: 697  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; OTHER INFORMATION: Genbank Accession No. US20030134324A1 W73859  
 ; NAME/KEY: unsure  
 ; LOCATION: (1)..(697)  
 ; OTHER INFORMATION: n = a o r c o r g o r t  
 US-09-873-319-679

Alignment Scores:  
 Pred. No.: 3,54e-30 Length: 697  
 Score: 318.00 Matches: 61  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 34.68% Indels: 0  
 DB: 10 Gaps: 0

US-09-701-674A-23 (1-179) x US-09-873-319-679 (1-697)

QY 119 AspThrLeuArgLeuAlaSerSerTyrIleAlaHisLeuArgGlnIleLeuAlaAsnAsp 138  
 Db 3 GACACGCTCAGGTCGGTCAGCTACATCGCCACTTGAGCGAGATCCTGGCTAACGAC 62  
 QY 139 LysTyrGluAsnGlyTyrIleHisProValAsnLeuThrTrpProPheMetValAlaGly 158  
 Db 63 AAATACGAGAACGGGTACATTACCCGGTCAACCTGAGCTGGCCCTTTATGTGGCCGGG 122  
 QY 159 LysProGluSerAspLeuLysGluValThrAlaSerArgLeuCysGlyThrThrAla 178  
 Db 123 AAACCCGAGAGTGACCTGAAAGAGTGGTGACCGGAGCGCGCTTATGTGGAACCCGCG 182  
 QY 179 Ser 179  
 Db 183 TCC 185

## RESULT 13

US-09-918-995-11181  
 ; Sequence 11181, Application US/09918995  
 ; Publication No. US20030073623A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Hyseq, Inc.  
 ; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED  
 ; FILE REFERENCE: 20411-756  
 ; CURRENT APPLICATION NUMBER: US/09/918,995  
 ; CURRENT FILING DATE: 2001-07-30  
 ; PRIOR APPLICATION NUMBER: US/09/235,076  
 ; PRIOR FILING DATE: 1999-01-20  
 ; NUMBER OF SEQ ID NOS: 38054  
 ; SOFTWARE: FastSeq for Windows Version 3.0  
 ; SEQ ID NO 11181  
 ; LENGTH: 493  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; NAME/KEY: misc\_feature



us-09-701-674a-23.rnpb

Fri Dec 17 09:25:28 2004

Job time : 491 secs



**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: December 16, 2004, 21:48:16 ; Search time 2870 Seconds

(without alignments)  
2272.722 Million cell updates/sec

Title: US-09-701-674A-23

Perfect score: 917

Sequence: 1 MTGSLSDVEDIQEVMLEEC.....PESDLKEVTVASRLCGTTAS 179

Scoring table: BLOSUM62

Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 32822875 seqs, 18219865908 residues

Total number of hits satisfying chosen parameters: 65645750

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-O=/cgp2\_1/USPTO.spool\_h/US09701674/runat\_15122004\_100441\_27517/app\_query.fasta\_1.327  
-DB=EST -QFMT=fastap -SUFFIX=est -MINMATCH=0.1 -LOOPEL=0 -LOOPEXI=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09701674 @CGN 1.1 3437 @runat\_15122004\_100441\_27517 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6  
-FGAPEXT=7 -XGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST:  
1: gb\_est1:  
2: gb\_est2:  
3: gb\_hc:  
4: gb\_est3:  
5: gb\_est4:  
6: gb\_est5:  
7: gb\_est6:  
8: gb\_gssi:  
9: gb\_gss2:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	917	100.0	782	6	CF128484
2	917	100.0	796	5	EX107489
3	917	100.0	1044	1	AL550251
4	917	100.0	1221	3	CR594669
5	917	100.0	1250	3	CR606057
6	917	100.0	1252	3	CR616308
7	912	99.5	1038	4	BM544256
8	911	99.3	1024	1	AL575799
9	910	99.2	1099	1	AL552119

10	909	99.1	800	6	CF127519
11	901	98.3	827	5	EX926295
12	901	98.3	1098	4	BM923344
13	900	98.1	1056	5	EX337699
14	890	97.1	1085	1	AL574170
15	887	96.7	1012	6	BY710826
16	887	96.7	1238	3	AK011575
17	883	96.3	996	5	EX360244
18	876	95.5	778	7	CO564290
19	876	95.5	965	1	AL552380
20	874	95.3	794	6	CF127034
21	870	94.9	726	5	EX920031
22	867	94.5	1169	1	AL552084
23	859	93.7	647	7	CF359503
24	855	93.2	712	5	EX922447
25	852	92.9	926	4	BI769090
26	851	92.8	563	1	AV605533
27	847	92.4	543	2	BF198267
28	845.5	92.2	1065	5	BQ072411
29	839	91.5	740	5	BU290232
30	839	91.5	879	5	BU459258
31	836	91.2	818	4	BI767574
32	827	90.2	597	5	BU358809
33	821	89.5	746	4	BI770998
34	820	89.4	785	6	CF127896
35	810	88.3	715	7	CK835890
36	810	88.3	718	7	CN409630
37	800	87.2	757	6	CF126927
38	799	87.1	675	5	EX919517
39	799	87.1	792	6	CF270487
40	798	87.0	645	6	CF128246
41	797	86.9	589	6	CB417689
42	797	86.9	744	6	CF128914
43	783.5	85.4	825	4	BI768711
44	781	85.2	717	6	CF126467
45	781	85.2	837	4	BI770596

#### ALIGNMENTS

CF128484 782 bp mRNA linear EST 05-AUG-2003  
UI-HF-ETO-awj-c-15-0-UI.r1 NIH MGC\_214 Homo sapiens CDNA clone  
IMAGE:30557534 5', mRNA sequence.

CF128484  
CF128484.1 GI:33207776

EST.  
Homo sapiens (human)

ORGANISM

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 782)

Bonaldo,M.F., Lennon,G. and Soares,M.B.

Normalization and subtraction: two approaches to facilitate gene

discovery

Genome Res. 6 (9), 791-806 (1996)

97044477

8889548

Contact: Soares, MB

Coordinated Laboratory for Computational Genomics

University of Iowa

375 Newton Road , 4156 MEBRF, Iowa City, IA 52242, USA

Tel: 319 335 8250

Fax: 319 335 9565

Email: bento-soares@uiowa.edu

Tissue Procurement: Mary Hendrix

cDNA Library preparation: Dr. M. Bento Soares, University of Iowa

cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa

Clone Sequencing by: Dr. M. Bento Soares, University of Iowa

Clone Distribution: Distribution information can be found at

http://genome.uiowa.edu/distribution/humanfl.html

The following repetitive elements were found in this cDNA

sequence: 1-81, >(GA)n#Simple\_repeat (matched complement)

Seq primer: pYX-5.

# FEATURES

Location/Qualifiers  
1..782  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:30557534"  
/tissue type="Chondrosarcoma Lung Metastasis cell lines"  
/lab host="DH10B (T1 phage resistant)"  
/clone lib="NH1 MGC 214"  
/notes="Organ: Lung; Vector: pYX-Asc; Site 1: EcoR I; Site 2: Not I; The library was constructed according to Bonaldi, Lennon and Soares, Genome Research, 6:791-806, 1996. Denatured RNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with oligo-dT primer containing a Not I site. Double strand cDNA was size selected according to mRNA size fraction, ligated with EcoR I adaptor, digested with Not I and then cloned directionally into pYX-Asc vector. The library tag sequence located between the Not I site and the polyA tail is GATAGGCCA. Tissue was provided by Mary Hendrix."

## ORIGIN

Alignment Scores:  
Pred. No.: 1,34e-86 Length: 782  
Score: 917.00 Matches: 179  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 6 Gaps: 0

US-09-701-674A-23 (1-179) x CF128484 (1-782)

Qy 1 MetSerThrGlySerLeuSerAspValGluAspLeuGlnGluValGluMetLeuGluCys 20  
Db 86 ATGTCACCGCGTCCCTCAGCGATGTGGAGGACCTTCAAGAGGTGGAGATGTGGAATGT 145  
Qy 21 AspGlyLeuLeuMetAspSerAsnLysGluPheValThrSerAsnGluSerThrGluGlu 40  
Db 146 GACGGTGTGAATGATTCGACACAGGAATTTGTGACTTCCACAGAGACACCGAGAG 205  
Qy 41 SerSerAsnGlyLeuAsnGlySerProGlnLysGlyArgGlyGlyLeuLysArgArg 60  
Db 206 AGCTCCAACTCGAGAATGGTCTCCCAAGAGGCGCGCGGCTGGCGCAAGAGGAGG 265  
Qy 61 LysAlaProThrLysLysSerProLeuSerGlyValSerGlnGluGlyGlnValGln 80  
Db 266 AAGGCGCCCAACCAAGAGGCGCCCTTGAGCGGGGTGAGCGAGGAGGAGGAGGAGG 325  
Qy 81 ArgAsnAlaAlaAsnAlaArgGluArgAlaArgMetArgValLeuSerLysAlaPheSer 100  
Db 326 CGCAACGCGCGCAACGCGGAGAGCGGCGCGCATCGAGTCTGAGCAAGCGCTTCTCC 385  
Qy 101 ArgLeuLysThrThrLeuProThrValProProAspThrLysLeuSerLysLeuAspThr 120  
Db 386 AGACTCAAGACCACTTCCCTGCGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 445  
Qy 121 LeuArgLeuAlaSerSerThrLeuAlaHisLeuArgGlnIleLeuAlaAsnAspLysTyr 140  
Db 446 CTGAGGTGGCGTCCAGCTACATCGCCCACTTGAGGCGAGATCTGCTAACACAAATAC 505  
Qy 141 GluAsnGlyTyrIleHisProValAsnLeuThrTrpPheMetValAlaGlyLysPro 160  
Db 506 GAGAACGGGTACATTCACCGGTCAACCTGACGTGGCGCTTTATGTGGTGGCGGAAACCC 565  
Qy 161 GluSerAspLeuLysGluValValThrAlaSerArgLeuCysGlyThrThrAlaSer 179  
Db 566 GAGAGTGACCTGAAGAGTGTGACCGGAGCGCGCTTATGTGGAAACCAACCGCGTCC 622

RESULT 2  
BX107489

LOCUS BX107489 796 bp mRNA linear EST 06-FEB-2003

## DEFINITION

EX107489 Soares Placenta Nb2HP Homo sapiens cDNA clone  
IMAGE998N24234 ; IMAGE:151511, mRNA sequence.

## ACCESSION

EX107489

## VERSION

EX107489.1 GI:27834707

## KEYWORDS

EST.

## SOURCE

Homo sapiens (human)

## ORGANISM

Homo sapiens (human)

## REFERENCE

1 (bases 1 to 796)

## AUTHORS

Ebert, L., Heil, O., Hennig, S., Neubert, P., Partsch, E., Peters, M., Radelof, U., Schneider, D. and Korn, B.

## TITLE

Human Unigeneset - RZPD3

## JOURNAL

Unpublished (2003)

## COMMENT

Contact: Ina Rolfs  
RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH  
Im Neuenheimer Feld 580, D-69120 Heidelberg, Germany  
RZPD; IMAGE998N24234.  
RZPDLIB; I.M.A.G.E. cDNA Clone Collection;  
Human Unigeneset - RZPD3 (RZPDLIB No.972)  
http://www.rzpd.de/CloneCards/cgi-bin/showlib.pl.cgi/response?libNo=972 Contact: Ina Rolfs  
RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH  
Heubnerweg 6, D-14059 Berlin, Germany  
Tel: +49 30 32639 101  
Fax: +49 30 32639 111  
www.rzpd.de

This clone is available royalty-free from RZPD;  
contact RZPD (clone@rzpd.de) for further information. Seq primer:  
M13r, Primer sequence: TTTCACAGGAAACAGCTATGAC.

## FEATURES

### Location/Qualifiers

1..796  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE998N24234 ; IMAGE:151511"  
/sex="female"  
/dev\_stage="placenta obtained at birth (full term)"  
/lab host="DH10B (ampicillin resistant)"  
/clone lib="Soares placenta Nb2HP"  
/notes="Organ: placenta; Vector: pT7T3D (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' AACTGAGAATTCGCGCGGAGGAAATTTTCTTTTCTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Fatima Bonaldi."

## ORIGIN

### Alignment Scores:

Pred. No.: 1,37e-86 Length: 796  
Score: 917.00 Matches: 179  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 5 Gaps: 0

US-09-701-674A-23 (1-179) x BX107489 (1-796)

Qy 1 MetSerThrGlySerLeuSerAspValGluAspLeuGlnGluValGluMetLeuGluCys 20

Db 195 ATGTCCACCGCGTCCCTCAGCGATGTGGAGGACCTTCAAGAGGTGGAGATGTGGATGT 254

Qy 21 AspGlyLeuLysMetAspSerAsnLysGluPheValThrSerAsnGluSerThrGluGlu 40

Db 255 GACGGGTGAATGATTCGAAACAGGAATTTGTGACTTCCACAGAGACCCGAGGAG 314

Qy 41 SerSerAsnGlyGluAsnGlySerProGlnLysGlyArgGlyGlyLeuGlyLysArgArg 60

Db 315 AGTCCCAACTGCGAGAAATGGGTCTCCCAAGAGGCGCGCGCGCTTGGCCAAAGAGGAG 374



```

/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0D1039YJ05"
/tissue_type="Placenta Cot 25-normalized"
/plasmid="PCMVSPORT_6"

ORIGIN
Alignment Scores:      2,45e-86      Length:      1221
Pred. No.:            917.00      Matches:      179
Score:                100.00%      Conservative: 0
Percent Similarity:    100.00%      Mismatches:  0
Best Local Similarity: 100.00%      Indels:      0
Query Match:          100.00%      Gaps:        3
DB:
3

US-09-701-674A-23 (1-179) x CR594669 (1-1221)
QY 1 MetSerThrGlySerLeuSerAspValGluAspLeuGlnGluValGluMetLeuGluCys 20
DB 260 ATGTCACCGCTCCCTCAGCATGTGGAGGACCTTCAAGAGTGGAGATGTTGAATGT 319
QY 21 AspGlyLeuLysMetAspSerAsnLysGluPheValThrSerAsnGluSerThrGluGlu 40
DB 320 GACGGGTGAAATGGAATTCGAACAAGGAATTTGTGACTTCCACGAGAGACCGGAGAG 379
QY 41 SerSerAsnGlySerProGlnLysGlyArgGlyGlyLeuGlyLysArg 60
DB 380 AGCTCCAACTCGAGATGGGCTCCCCAGAAAGGCGCGGCTGGGCAAGAGGAGG 439
QY 61 LysAlaProThrLysLysSerProLeuSerGlyValSerGlnGluGlyLysGlnValGln 80
DB 440 AAGGCGCCACCAAGAAGAGCCCTCAGCGGGGTGAGCAGGAGGAGGAGGAGTCCAG 499
QY 81 ArgAsnAlaAlaSerThrLeuGluArgAlaArgMetArgValLeuSerLysAlaPheSer 100
DB 500 CGCAACCGCCCAACCGCGAGAGCGGCGCCGCGATGCGAGTGTCTGAGCAAGGCTTCTCC 559
QY 101 ArgLeuLysThrThrLeuProThrValProProAspThrLysLeuSerLysLeuAspThr 120
DB 560 AGACTCAAGACCACTGCGCTGGGTGGTGGCGCCCGACACCAAGCTCTTCCAAAGTGGACACG 619
QY 121 LeuArgLeuAlaSerSerThrLeuAlaHisLeuArgGlnLeuAlaAsnAspLysTyr 140
DB 620 CTGAGGTGGCGTCCAGTACATCGCCCACTTGAGCAGATCTGCTTACGACCAAAATAC 679
QY 141 GluAsnGlyTyrIleHisProValAsnLeuThrTrpProPheMetValAlaGlyLysPro 160
DB 680 GAGACGGGTACATTCACCGGTCAACCTGACGTGGCCCTTTATGTGGCCCGGGAACCC 739
QY 161 GluSerAspLeuLysGluValThrAlaSerArgLeuCysGlyThrThrAlaSer 179
DB 740 GAGAGTGACCTGAAGAAGTGGTGCAGCGGAGCGGCTTATGTGGAACCAACCGCGTCC 796

RESULT 5
CR606057 1250 bp mRNA linear HTC 21-JUL-2004
LOCUS full-length cDNA clone CS0D1039YJ05 of Placenta Cot 25-normalized
DEFINITION of Homo sapiens (human).
ACCESSION CR606057
VERSION CR606057.1 GI:50486864
KEYWORDS HTC; CNSLT cDNA.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1250)
Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished
Contact : Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Faraday Avenue
2 (bases 1 to 1250)

REFERENCE
AUTHORS CR616308 1252 bp mRNA linear HTC 21-JUL-2004
TITLE full-length cDNA clone CS0D1070YK16 of Placenta Cot 25-normalized
JOURNAL of Homo sapiens (human).
REMARK CR616308
CR616308.1 GI:50497115

```

```

Genoscope.
Direct Submission
Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :
BP 131 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)
- Web : www.genoscope.cns.fr)
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-strand cDNA was digested with Not I and cloned
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
was normalized. Library was constructed by Life Technologies, a
division of Invitrogen.
FEATURES
Location/Qualifiers
1..1250
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0D1039YJ05"
/tissue_type="Placenta Cot 25-normalized"
/plasmid="PCMVSPORT_6"

ORIGIN
Alignment Scores:      2,53e-86      Length:      1250
Pred. No.:            917.00      Matches:      179
Score:                100.00%      Conservative: 0
Percent Similarity:    100.00%      Mismatches:  0
Best Local Similarity: 100.00%      Indels:      0
Query Match:          100.00%      Gaps:        3
DB:
3

US-09-701-674A-23 (1-179) x CR606057 (1-1250)
QY 1 MetSerThrGlySerLeuSerAspValGluAspLeuGlnGluValGluMetLeuGluCys 20
DB 278 ATGTCCACCGCTCCCTCAGCATGTGGAGGACCTTCAAGAGTGGAGATGTTGAATGT 337
QY 21 AspGlyLeuLysMetAspSerAsnLysGluPheValThrSerAsnGluSerThrGluGlu 40
DB 338 GACGGGTGAAATGGAATTCGAACAAGGAATTTGTGACTTCCAAACGAGAGACCGGAGG 397
QY 41 SerSerAsnGlySerProGlnLysGlyArgGlyGlyLeuGlyLysArg 60
DB 398 AGTCCAACTGCGAGATGGGTCTCCCGAGAAAGGCGCGGCTGGGCAAGAGAGAG 457
QY 61 LysAlaProThrLysLysSerProLeuSerGlyValSerGlnGluGlyLysGlnValGln 80
DB 458 AAGCGCGCCACCAAGAAGAGACCCCTCAGCGGGTCCAGCCAGGAGGAGGAGTCCAG 517
QY 81 ArgAsnAlaAlaSerThrLeuGluArgAlaArgMetArgValLeuSerLysAlaPheSer 100
DB 518 CGCAACCGCCCAACCGCGAGAGCGGCGCCGATCGAGTGTGAGCAAGGCTTCTTCC 577
QY 101 ArgLeuLysThrThrLeuProThrValProProAspThrLysLeuSerLysLeuAspThr 120
DB 578 AGACTCAAGACCACTGCGCTGGGTGGCGCCCGACACCAAGCTCTTCCAGCTGGACACG 637
QY 121 LeuArgLeuAlaSerSerThrIleAlaHisLeuArgGlnLeuAlaAsnAspLysTyr 140
DB 638 CTCAGGTGGGTGCTCAGTACATCGCCCACTTGGAGGAGATCTTGGCTAACCAAAATAC 697
QY 141 GluAsnGlyTyrIleHisProValAsnLeuThrTrpProPheMetValAlaGlyLysPro 160
DB 698 GAGAACGGGTACATTCACCGGTCAACCTGACGTGGCCCTTTATGTGGCCCGGGAACCC 757
QY 161 GluSerAspLeuLysGluValThrAlaSerArgLeuCysGlyThrThrAlaSer 179
DB 758 GAGAGTGACCTGAAGAAGTGGTGCAGCGGAGCGGCTTATGTGGAACCAACCGCGTCC 814

RESULT 6
CR616308 1252 bp mRNA linear HTC 21-JUL-2004
LOCUS full-length cDNA clone CS0D1070YK16 of Placenta Cot 25-normalized
DEFINITION of Homo sapiens (human).
ACCESSION CR616308
VERSION CR616308.1 GI:50497115

```

```

KEYWORDS HTC: CNSLT_CDNA.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
JOURNAL 1 (bases 1 to 1252)
REMARK Full-length cDNA libraries and normalization
Contact : feng liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Faraday Avenue
Genoscope.
2 (bases 1 to 1252)
Direct Submission
Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :
BP 131 91006 EVRY cedex - FRANCE [E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr]
COMMENT 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-strand cDNA was digested with Not I and cloned
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
was normalized. Library was constructed by Life Technologies, a
division of Invitrogen.
FEATURES
Location/Qualifiers
1..1252
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS01070YK16"
/tissue_type="Placenta"
/plasmid="pCMVSPORT_6"
ORIGIN
Alignment Scores:
Pred. No.: 2 54e-86 Length: 1252
Score: 917.00 Matches: 179
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 3 Gaps: 0
US-09-701-674A-23 (1-179) x CH616308 (1-1252)
QY 1 MetSerThrGlySerLeuSerAspValGluApsLeuGlnGluValGluMetLeuGluCys 20
272 ARGTCACCGCGTCCCTCAGCGATGTGGAGGACCTTCAAGAGTGAGATGTTGGAATGT 331
QY 21 AspGlyLeuLysMetAspSerAsnLysGluPheValThrSerAsnGluSerThrGluGlu 40
332 GACGGTTGAAATGATTCGACACAGGAATTTGTGACTTCCACGAGACCCGAGGAG 391
QY 41 SerSerAsnGlySerProGlnLysGlyArgGlyLeuGlyLysArgArg 60
392 AGTCCCACTGCGAGAAATGGGTCTCCCAAGAGGGCGCGCGGCTGGGCAAGAGGAGG 451
QY 61 LysAlaProThrLysLysSerProLeuSerGlyValSerGlnGluGlyValGln 80
452 AAGGCCCCACCAAGAGAGGCCCTTACGGGGGTACGACGAGGGGGAAGCAGGTCCAG 511
QY 81 ArgAsnAlaAlaAsnAlaArgGluArgAlaArgMetArgValLeuSerLysAlaPheSer 100
512 CGCAACGCCGCCAACCGCGAGAGCGGCCCGCATGGCGAGTCTGAGCAAGGCCTTCTCC 571
QY 101 ArgLeuLysThrLeuProTrpValProProAspThrLysLeuSerLysLeuAspThr 120
572 AGACTCAAGACCAACCCCTGGCCCTGGGTGCCCCCGACCAACCAAGCTCTCCAAGTGGACAG 631
QY 121 LeuArgLeuAlaSerSerTyrIleAlaHisLeuArgGlnIleLeuAlaAsnAspLysTyr 140
632 CTCAGGCTGGCGTCCAGCTACATCCCCACTTGGAGCGAGATCTCTGGCTTACCAACATAC 691
QY 141 GluAsnGlyTyrIleHisProValAsnLeuThrTrpProPheMetValAlaGlyLysPro 160

```

---

```

Db 692 GAGAACGGGTACATTCACCCCGTCAACCTGACGTGGCGCCCTTTATGGTGGCGGAAACCC 751
QY 161 GluSerAspLeuLysGluValValThrAlaSerArgLeuCysGlyThrThrAlaSer 179
Db 752 GAGAGTGACCTGAAAGAGAGTGCTGACCGGAGCGCTTATGTGGAACCCACCGCTCC 808
RESULT 7
BM544256
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1038)
NIH-MGC http://mgc.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Invitrogen
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM12357 row: m column: 09
High quality sequence stop: 774.
FEATURES
Location/Qualifiers
1..1038
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5587760"
/lab_host="DH108"
/clone_lib="NIH_MGC_125"
/notes="Organ: ovary (pool of 3); Vector: pCMV-SPORT6;
Site1: EcoRV (destroyed); Site2: NotI; RNA source pool
of three ovaries, from females ranging in age from 38 to
49 yo. Library is oligo-dT primed and directionally cloned
(EcoRV site is destroyed upon cloning). Average insert
size 2.1 kb, insert size range 1-3.5 kb. Library is
normalized and enriched for full-length clones and was
constructed by C. Gruber (Invitrogen). Research Genetics
tracking code 036."
ORIGIN
Alignment Scores:
Pred. No.: 6 65e-86 Length: 1038
Score: 912.00 Matches: 178
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 99.45% Indels: 0
DB: 4 Gaps: 0
US-09-701-674A-23 (1-179) x BM544256 (1-1038)
QY 2 SerThrGlySerLeuSerAspValGluApsLeuGlnGluValGluMetLeuGluCysAsp 21
82 TCCACCGCGTCCCTCAGCGATGTGGAGGACCTTCAAGAGTGAGATGTTGGAATGTGAC 141
QY 22 GlyLeuLysMetAspSerAsnLysGluPheValThrSerAsnGluSerThrGluGluSer 41
142 GGGTTGAAATGATTCGACACAGGAATTTGTGACTTCCACGAGACCCGAGGAGAGC 201
QY 42 SerAsnGlyLeuAsnGlySerProGlnLysGlyArgGlyLeuGlyLysArgGlyLys 61
202 TCCAACTCGAGAAATGGGTCTCCCAAGAGGGCGCGGCTGGCAAGAGGAGGAGAG 261

```



```

/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0D1059YH1.7"
/tissue_type="PLACENTA COT 25-NORMALIZED"
/clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
/notes="1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."

```

## ORIGIN

```

Alignment Scores:
  Length: 1178-85
  Score: 910.00
  Percent Similarity: 99.44%
  Best Local Similarity: 99.44%
  Query Match: 99.24%
  DB: 1
  Gaps: 0
  Indels: 0
  Mismatches: 0
  Conservatives: 178
  Matches: 1099

US-09-701-674A-23 (1-179) x AL552119 (1-1099)

QY 1 MetSerThrGlySerLeuSerAspValGluAspLeuGlnGluValGluMetLeuGluCys 20
Db 260 ATGTCCACCGCTCCCTCAGCGATGTGGAGGACCTTCAGAGGTGGAGATGTTGGAATGT 319
QY 21 AspGlyLeuLysMetAspSerLeuLysGluPheValThrSerAsnGluSerThrGluGlu 40
Db 320 GACGGGTGAAATGGAATTCGAACAAGGAATTTGTGACTTCCACGAGAGCACCAGGAG 379
QY 41 SerSerAsnGlySerLeuGlnGlySerProGlnLysGlyArgGlyGlyLeuGlyLysArg 60
Db 380 AGCTCCAAATGGAGAAATGGTCTCCCAAGAGGGCGCGCGGCTGGCGCAAGAGGAGG 439
QY 61 LysAlaProThrLysLysSerProLeuSerGlyValSerGlnGluGlyLysGlnValGln 80
Db 440 AAGCGCCCAACAGAGAGCGCCCTGAGCGGGTCAGCCAGAGGGAGAGCAGGTCCAG 499
QY 81 ArgAsnAlaAlaAsnAlaArgGluArgAlaArgMetArgValLeuSerLysAlaPheSer 100
Db 500 CGCAAGCGCCCAACAGAGAGGGCGCGCGGCTGCGAGTCTCAGCAAGCGCTTCCTCC 559
QY 101 ArgLeuLysThrThrLeuProTrpValProAspThrLysLeuSerLysLeuAspThr 120
Db 560 AGACTCAAGACCACTGCTGCTGGTGCCCGCCGACCAAGCTTCCAGCTGGACAG 619
QY 121 LeuArgLeuAlaSerSerTyrIleAlaHisLeuArgGlnIleLeuAlaAsnAspLysTyr 140
Db 620 CTCAGCTGGCTCCAGCTACATCGCCCACTTGAGGCGAGATCCTGGCTAAACGACAAATAC 679
QY 141 GluAsnGlyTyrIleHisProValAsnLeuThrTrpProPheMetValAlaGlyLysPro 160
Db 680 GAGACGGGTATATTACCCCGGTCAACCTGACGTGGCCCTTTATGGTGGCGGGAACCC 739
QY 161 GluSerAspLeuLysGluValValThrAlaSerArgLeuLysGlyThrThrAlaSer 179
Db 740 GAGAGTKACCTGAAGAAGTGGTGACCGCGAGCGCGCTTATGTGGAACCAACCGCGTCC 796

```

## RESULT 10

```

CF127519
LOCUS CF127519
DEFINITION UI-HF-ET0-awh-f-14-0-UI.r1 NIH MGC_214 Homo sapiens cDNA clone
IMAGE:30554533 5', mRNA sequence.
CF127519
CF127519.1 GI:33205841
EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 800)
Bonaldo, M.F., Lennon, G. and Soares, M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery

```

```

JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
PUBMED 889548
COMMENT

```

Contact: Soares, MB  
Coordinated Laboratory for Computational Genomics  
University of Iowa  
375 Newton Road, 4156 MBRF, Iowa City, IA 52242, USA  
Tel: 319 335 8250  
Fax: 319 335 9585  
Email: bento-soares@uiowa.edu

Tissue Procurement: Mary Hendrix  
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa  
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa

DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
Clone Distribution: Distribution information can be found at  
<http://genome.uiowa.edu/distribution/humanfl.html>  
The following repetitive elements were found in this cDNA  
sequence: 172-253, >(GA)n#simple\_repeat (matched complement)  
Seq primer: pYX-5.

## FEATURES

Location/Qualifiers

Source

```

1..800
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:30554533"
/tissue_type="Chondrosarcoma Lung Metastasis cell lines"
/lab_host="DH10B (T1 phage resistant)"
/clone_lib="NIH MGC_214"
/notes="Organ: Lung; Vector: pYX-Asc; Site 1: EcoR I; Site 2: Not I; The library was constructed according Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. Denatured RNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with oligo-dT primer containing a Not I site. Double strand cDNA was size selected according to mRNA size fraction, ligated with EcoR I adaptor, digested with Not I and then cloned directionally into pYX-Asc vector. The library tag sequence located between the Not I site and the polyA tail is GATAAGGCCA. Tissue was provided by Mary Hendrix."

```

## ORIGIN

```

Alignment Scores:
  Length: 800
  Score: 909.00
  Percent Similarity: 100.00%
  Best Local Similarity: 100.00%
  Query Match: 99.13%
  DB: 6
  Gaps: 0
  Indels: 0
  Mismatches: 0
  Conservatives: 0
  Matches: 177

US-09-701-674A-23 (1-179) x CF127519 (1-800)

QY 1 MetSerThrGlySerLeuSerAspValGluAspLeuGlnGluValGluMetLeuGluCys 20
Db 258 ATGTCCACCGCTCCCTCAGCGATGTGGAGGACCTTCAGAGGTGGAGATGTTGGAATGT 317
QY 21 AspGlyLeuLysMetAspSerAsnLysGluPheValThrSerAsnGluSerThrGluGlu 40
Db 318 GACGGGTGAAATGGAATTCGAACAAGGAATTTGTGACTTCCACGAGAGCACCAGGAG 377
QY 41 SerSerAsnGlySerLeuGlnGlySerProGlnLysGlyArgGlyGlyLeuGlyLysArg 60
Db 378 AGCTCCAAATGGAGAAATGGTCTCCCAAGAGGGCGCGCGGCTGGCGCAAGAGGAGG 437
QY 61 LysAlaProThrLysLysSerProLeuSerGlyValSerGlnGluGlyLysGlnValGln 80
Db 438 AAGCGCCCAACAGAGAGCGCCCTGAGCGGGTTCAGCGAGGAGGAGGAGGAGGAGGAG 497
QY 81 ArgAsnAlaAlaAsnAlaArgGluArgAlaArgMetArgValLeuSerLysAlaPheSer 100
Db 498 CGCAAGCGCCCAACAGAGAGCGCGGCGCGGCGCGGCTGCTGAGCAGAGCGCTTCCTCC 557
QY 101 ArgLeuLysThrThrLeuProTrpValProAspThrLysLeuSerLysLeuAspThr 120

```



```

Db      558 AGACTCAAGACACCCCTGCCCTGGTGGTGGCCCGCCGACCAACAGCTTCCAAAGCTGGACAGC 617
QY      121 LeuArgLeuAlaSerSerTyrTyrTyrTyrTyrTyrTyrTyrTyrTyrTyrTyrTyrTyrTyr 140
Db      618 CTCAGGCTGGCTGCAGCTACATCCGCCACCTTGAGGCGAGATCCCTGGCTAACGACAAATAC 677
QY      141 GluAsnGlyTyrTyrTyrTyrTyrTyrTyrTyrTyrTyrTyrTyrTyrTyrTyrTyrTyrTyr 160
Db      678 GAGAACGGGTACATTCAACCTGGTCAACCTGAGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCT 737
QY      161 GluSerAspLeuGluValValThrAlaSerArgLeuCysGlyThrThr 177
Db      738 GAGAGTGAGCTGAAGAAGTGGTGAGCCGCGAGCCCTTATGTGGAAACACG 788

RESULT 11
LOCUS   BX926295
DEFINITION BX926295 Sus Scrofa library (scan) Sus scrofa cDNA clone
ACCESSION scna0021d.a.24 5prim, mRNA sequence.
VERSION   BX926295
KEYWORDS  BX926295.1 GI:41143143
SOURCE    EST.
ORGANISM  Sus scrofa (pig)
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE 1 (bases 1 to 827)
AUTHORS   Bonnet,A., Tosser-Klopp,G., Benne,F., Cabau,C., Villegier,S.,
          Soares,M., Bonaldo,F. and Hately,F.
          A Pig Normalised Multi-Tissue cDNA Library
          Unpublished (2003)
          Contact: Tosser-Klopp G
          Genetique Animale
          Institut National de la Recherche Agronomique
          Chemin de Borde-Rouge - Auzeville BP27, 31326 Castanet-Tolosan
          cedex, FRANCE
          Tel: 33 (0) 5.61.28.51.14
          Fax: 33 (0) 5.61.28.53.08
          Email: tosser@toulouse.inra.fr
          Sequence cleaned of vector, adaptor and repetitions. Contact us
          at sigenasupport@jouy.inra.fr to obtain the chromatogram of this
          sequence.
          Plate: 0021 row: a column: 24.
          Location/Qualifiers
            1..827
              /organism="Sus scrofa"
              /mol_type="mRNA"
              /db_xref="taxon:9823"
              /clone="scan0021d.a.24"
              /tissue_type="mixed"
              /dev_stage="from embryos to adults"
              /clone_lib="Sus Scrofa library (scan)"
              /note="Tissues: adipose tissue, brain, kidney, liver,
              muscle, ovary, testis, heart, hypothalamus, pancreas,
              skin, spleen, thymus, placenta, pituitary gland, seminal
              vesicle, small intestine, uterus, adrenals, bulbo urethral
              gland, cerebral trunk, epididymis, female gonad,
              gall-bladder, hippocampus, large intestine, male gonad,
              melanocytes, stomach, udder"

ORIGIN
Alignment Scores:
Pred. No.: 7,12e-85 Length: 827
Score: 901.00 Matches: 176
Percent Similarity: 98.88% Conservative: 1
Best Local Similarity: 98.32% Mismatches: 2
Query Match: 98.26% Indels: 0
DB: 5 Gaps: 0

US-09-701-674A-23 (1-179) x BX926295 (1-827)
QY      1 MetSerThrGlySerLeuSerAspValGluAspLeuGlnGluValGluMetLeuGluCys 20

```

```

Db      192 ATGTCCACCGGCTCCCTCAGCGATGTGGAGGACCTTCAGAGGTTGGAGATGCTGGACTGC 251
QY      21 AspGlyLeuLysMetAspSerAsnLysGluPheValThrSerAsnGluSerThrGluGlu 40
Db      252 GACGGGCTGAAATGAGACTCGAACAGAGGAGTTTGTGACTTCCACGAGAGTACCGAGGAG 311
QY      41 SerSerAsnCysGluAsnGlySerProGlnLysGlyValArgGlyGlyLeuGlyLysArgArg 60
Db      312 AGTCTCAACTGCGTGCACCGGCTCTCCCCAGAAAGGCGCGCGGCTGGGCGGCAAGGAGG 371
QY      61 LysAlaProThrLysLysSerProLeuSerGlyValSerGlnGlnGlyLysGlnValGln 80
Db      372 AAGCGCCCCACCAAGAGAGGCCCTTTAGCGGGGTTCAGCCAGGAGGGAAGCAGGTCCAG 431
QY      81 ArgAsnAlaAlaAsnAlaArgGluArgAlaArgMetArgValLeuSerLysAlaPheSer 100
Db      432 CGCAACGCGGCCAACCGCGCGAGCGCGGCGGATCGGGGTCTGAGCAAGGCTTCTCC 491
QY      101 ArgLeuLysThrThrLeuProTyrValProProAspThrLysLeuSerLysLeuAspThr 120
Db      492 AGGCTCAAGACCACTGCTCCCTGGGTGCCCGGACACCAAGCTCTCCAAGCTGGACACC 551
QY      121 LeuArgLeuAlaSerSerTyrTyrTyrTyrTyrTyrTyrTyrTyrTyrTyrTyrTyrTyrTyr 140
Db      552 CTCAGGCTGGCTGCCAGTACATCGCCCACTTAAGGCAGATCTTGCCCAACGACAGTAC 611
QY      141 GluAsnGlyTyrTyrTyrTyrTyrTyrTyrTyrTyrTyrTyrTyrTyrTyrTyrTyrTyrTyr 160
Db      612 GAGAAATGTTATATTACCGGTCACCTGACGTGGCGCCCTTTATGGTGGCGGGAACCC 671
QY      161 GluSerAspLeuLysGluValValThrAlaSerArgLeuCysGlyThrThrAlaSer 179
Db      672 GAAAGTGACCTGAAGAAGTGGTAACTGGGAGCGGCTTTATGTGGAAACCAACGCGCTCC 728

BM923344 1098 bp mRNA linear EST 12-MAR-2002
AGENCOURT_6626079 NIH_MGC_116 Homo sapiens cDNA clone IMAGE:5758942
5', mRNA sequence.
BM923344
BM923344.1 GI:19373723
EST.
SOURCE    Homo sapiens (human)
ORGANISM  Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1098)
AUTHORS   NIH-MGC http://mgs.nci.nih.gov/.
TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT    Contact: Robert Strausberg, Ph.D.
          Email: cgapbs-@email.nih.gov
          Tissue Procurement: Life Technologies, Inc.
          cDNA Library Preparation: Life Technologies, Inc.
          cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
          DNA Sequencing by: Agencourt Bioscience Corporation
          Clone distribution: MGC clone distribution information can be
          found through the I.M.A.G.E. Consortium/LLNL at:
          http://image.llnl.gov
          Plate: LLAM12803 row: i column: 23
          High quality sequence stop: 693.
          Location/Qualifiers
            1..1098
              /organism="Homo sapiens"
              /mol_type="mRNA"
              /db_xref="taxon:9606"
              /clone="IMAGE:5758942"
              /lab_host="DH10B"
              /clone_lib="NIH_MGC_116"
              /note="Organ: pooled colon, kidney, stomach; Vector:
              pCMV-SF0R16; Site: 1: Not1; Site 2: EcoRV (destroyed); RNA
              source anonymous pool of 3 colons, age 26 yo male, 49 yo
              female, 71 yo male colon; 46 yo male kidney, and pool of 2

```

stomachs, 62 yo male and 70 yo female. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.4 kb, insert size range 1-3 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 023. Note: this is a NIH\_MGC Library."

## ORIGIN

## Alignment Scores:

Pred. No.: 1.05e-84 Length: 1098  
Score: 901.00 Matches: 178  
Percent Similarity: 99.44% Conservatives: 1  
Best Local Similarity: 98.89% Mismatches: 0  
Query Match: 98.26% Indels: 1  
DB: 4 Gaps: 0

US-09-701-674A-23 (1-179) x BM923344 (1-1098)

QY 1 MetSerThrGlySerLeuSerAspValGluAspLeuGlnGluValGluMetLeuGluCys 20  
Db 268 ATGTCACCGCGCTCCCTCAGCGATGTGGAGACCTTCAAGAGGTGGAGATGTTGAATGT 327  
QY 21 AspGlyLeuLysMetAspSerAsnLysGluPheValThrSerAsnGluSerThrGluGlu 40  
Db 328 GACGGGTGGAATGATTCGACACAGGAATTTGTGACTTCCACGAGACCGAGGAG 387  
QY 41 SerSerAsnCysGluAsnGlySerProGlnLysGlyArgGlyLeuGlyLysArg 60  
Db 388 AGCTCCAACTGCGAGAAATGGTGTCCCCAGAGGGCGCGGGCGCTGGCGAAGAGGAG 447  
QY 61 LysAlaProThrLysLysSerProLeuSerGlyValSerGlnGluLysGlnValGln 80  
Db 448 AAGCGCCCAACCAAGAAAGAGCCCTCAGCGGGTCCAGAGAGGGAGAGGATCCAG 507  
QY 81 ArgAsnAlaAlaAsnAlaArgGluArgAlaArgMetArgValLeuSerLysAlaPheSer 100  
Db 508 CGCACCGCGCCCAACCGCGAGAGCGCGCGCATCGAGTCTGAGCAAGGCCCTTCTCC 567  
QY 101 ArgLeuLysThrThrLeuProTrpValProAspThrLysLeuSerLysLeuAspThr 120  
Db 568 AGACTCAAGACCACTGCGCGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 627  
QY 121 LeuArgLeuAlaSerSerTyrIleAlaHisLeuArgGlnIleLeuAlaAsnAspLysTyr 140  
Db 628 CTCAGCGTGGGTCCAGCTACATCGCCCACTTGAGGCAGATCTGGCTAACGACAAATAC 687  
QY 141 GluAsnGlyTyrIleHisProValAsnLeuThrTrp-ProPheMetValAlaGlyLysPr 160  
Db 688 GAGAACGGGTACATTCACCGGTCAACCTGAGCTGGCCCTTTATGGTGGCGGAAACC 747  
QY 160 oGluSerAspLeuLysGluValValThrAlaSerArgLeuCysGlyThrThrAlaSer 179  
Db 748 CGAGAGTGACCTGGAAGAAATGTTGACCGAGAGCGCCCTTATGTGGAACCCCGCTCC 805

## RESULT 13

EX337699

LOCUS

DEFINITION EX337699 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA

clone CS0D1048YE15 5-PRIME, mRNA sequence.

ACCESSION EX337699

VERSION EX337699.2

KEYWORDS

SOURCE

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 1056)

Li, W.B., Gruber, C., Jessee, J. and Polayes, D.

Full-length cDNA libraries and normalization

Unpublished (2001)

On May 2, 2003 this sequence version replaced gi:30345636.

Contact: Genoscope

RESULT 14

AL574170/c

Genoscope - Centre National de Sequencage

BP 191 91006 EVRY cedex - France

Email: seqref@genoscope.cns.fr, web : www.genoscope.cns.fr

1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies, a division of Invitrogen. This sequence belongs to sequence cluster 3812.f

For more information about this cluster, see

http://www.genoscope.cns.fr/cdna?s=CS0D1048AC08QPl&c=3812.f.

## Location/Qualifiers

1..1056

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="CS0D1048YE15"

/tissue\_type="PLACENTA COT 25-NORMALIZED"

/clone\_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"

/primer="1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."

## ORIGIN

## Alignment Scores:

Pred. No.: 1.27e-84 Length: 1056  
Score: 900.00 Matches: 176  
Percent Similarity: 98.32% Conservatives: 0  
Best Local Similarity: 98.32% Mismatches: 3  
Query Match: 98.15% Indels: 0  
DB: 5 Gaps: 0

US-09-701-674A-23 (1-179) x BX337699 (1-1056)

QY 1 MetSerThrGlySerLeuSerAspValGluAspLeuGlnGluValGluMetLeuGluCys 20  
Db 278 ATGTCACCGCGCTCCCTCAGCGATGTGGAGACCTTCAAGAGGTGGAGATGTTGAATGT 337  
QY 21 AspGlyLeuLysMetAspSerAsnLysGluPheValThrSerAsnGluSerThrGluGlu 40  
Db 338 GACGGGTGAAATGGATTCGAAACAGGAATTTGTGACTTCCACAGAGACCGAGGAG 397  
QY 41 SerSerAsnCysGluAsnGlySerProGlnLysGlyArgGlyLeuGlyLysArg 60  
Db 398 AGCTCCAACTGCGAGAAATGGGTCTCCCAAGAGGGCGCGCGCTGGCGAAGAGAGG 457  
QY 61 LysAlaProThrLysLysSerProLeuSerGlyValSerGlnGluGlyLysGlnValGln 80  
Db 458 AAGCGCGCCCAAGAGAGCGCCCTGAGCGGGTCAAGCAGGAGGGGAAAGCAGGTCCAG 517  
QY 81 ArgAsnAlaAlaAsnAlaArgGluArgAlaArgMetArgValLeuSerLysAlaPheSer 100  
Db 518 CGCAACCGCCCAACCGCGAGAGGGGCGCGCATCGAGTGTGAGCAAGGCCCTTCTCC 577  
QY 101 ArgLeuLysThrThrLeuProTrpValProAspThrLysLeuSerLysLeuAspThr 120  
Db 578 AGACTCAAGACCACTCCCTGCGCTGGGTGCGCCCGGACACCAAGCTCCTCAAGCTGGACAG 637  
QY 121 LeuArgLeuAlaSerSerTyrIleAlaHisLeuArgGlnIleLeuAlaAsnAspLysTyr 140  
Db 638 CTCAGGTGGCGTCCAGTACATCGCCCACTTGAAGCAGATCCTGGCTAACGACAAATAC 697  
QY 141 GluAsnGlyTyrIleHisProValAsnLeuThrTrpProPheMetValAlaGlyLysPro 160  
Db 598 GAGAACGGGTACATTCACCGGTCAACCTGAGTGGCCCTTTATGGTGGCGGGAACCC 757  
QY 161 GluSerAspLeuLysGluValValThrAlaSerArgLeuCysGlyThrThrAlaSer 179  
Db 758 GAGAGTCACTGAAARAAGTGGTCAACCGAGCGCGCTTATGTGGAACCAACCGCGTCC 814

```

LOCUS       AL574170 Homo sapiens Placenta COT 25-NORMALIZED Homo sapiens cDNA
DEFINITION  clone CS0D1039YJ05 3-PRIME, mRNA sequence.
ACCESSION   AL574170
VERSION     AL574170.3 GI:46247266
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 1085)
AUTHORS     Li, W.-B., Gruber, C., Jessee, J. and Polayes, D.
TITLE       Full-length cDNA libraries and normalization
JOURNAL     Unpublished (2001)
COMMENT     On Feb 16, 2001 this sequence version replaced gi:31295505.
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: segre@genoscope.cns.fr, Web : www.genoscope.cns.fr
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-strand cDNA was digested with Not I and cloned
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
was normalized. Library was constructed by Life Technologies, a
division of Invitrogen. This sequence belongs to sequence cluster
3812.f
For more information about this cluster, see
http://www.genoscope.cns.fr/cdna?s=CS0D1039CE03NP1&c=3812.f.
FEATURES             source
     source
     1..1085
         /organism="Homo sapiens"
         /mol_type="mRNA"
         /db_xref="taxon:9606"
         /clone="CS0D1039YJ05"
         /tissue_type="PLACENTA COT 25-NORMALIZED"
         /clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
         /note="1st strand cDNA was primed with a NotI-oligo(dT)
         primer. Five prime end enriched, double-strand cDNA was
         digested with Not I and cloned into the Not I and EcoR V
         sites of the pCMVSPORT 6 vector. Library was normalized."
ORIGIN
Alignment Scores:
Pred. No.:      1,51e-83      Length:      1085
Score:          890.00      Matches:      178
Percent Similarity: 99.44%      Conservative: 0
Best Local Similarity: 99.44%      Mismatches: 1
Query Match:    97.06%      Indels:      1
DB:              1          Gaps:        0

US-09-701-674a-23 (1-179) x AL574170 (1-1085)
QY      1 MetSerThrGlySerLeuSerAspValGluAspLeuGlnGluValGluMetLeuGluCys 20
DB      969 ATGTCACCGGCTCCCTCAGCATGTGGAGGACCTTCAAGAGGTGGAGATGTGGAATGT 910
QY      21 AspGlyLeuLeuMetAspSerAnLysGluPheValThrSerAsnGluSerThrGluGlu 40
DB      909 GACGGGTGTAAGATGATTCGAACAGGAATTTGTGACTTCCACGAGACCCGAGGAG 850
QY      41 SerSerAsnCysGluAsnGlySerProGlnLysGlyArgGlyGlyLeuGlyLysArgArg 60
DB      849 AGCTCCAACTCGAGAATGGTCTCCAGAGAGCGCGCGCGCTGGCCAGAGAGGAGG 790
QY      61 LysAlaProThrLysLysSerProLeuSerGlyValSerGlnGluGlyLysGlnValGln 80
DB      789 AAGGCGCCCAACCAAGAGAGCCCTTGGAGGGGTTCAGCGAGGGGAGGAGGAGTCCAG 730
QY      81 ArgAsnAlaAlaAsnAlaArgGluArgAlaArgMetArgValLeuSerLysAlaPheSer 100
DB      729 CGCAACGCGCCCAACGCGGAGAGCGGGCCCGCATCGAGTGTGAGCAAGGCCCTTCCTC 670
QY      101 ArgLeuLysThrThrLeuProThrValProProAspThrLysLeuSerLysLeuAspThr 120

```

```

Db      669 AGACTCAAGACCAACCCCTGCCCTGGTG-CCCCCGCACCAACCAAGCTCTCCAGTGGACAG 611
QY      121 LeuArgLeuAlaSerSerTyrIleAlaHisLeuArgGlnIleLeuAlaAsnAspLysTyr 140
Db      610 CTCAGGCTGGCGTCCAGCTACATCGCCCACTTGGAGCAGATCTGGCTAACCAATATAC 551
QY      141 GluAsnGlyTyrIleHisProValAsnLeuThrTrpProPheMetValAlaGlyLysPro 160
Db      550 GAGAACGGGTACATTCAACCGGTCAACCTGACGTGGCCCTTTATGGTGGCGGAAACCC 491
QY      161 GluSerAspLeuLysGluValValThrAlaSerArgLeuCysGlyThrThrAlaSer 179
Db      490 GAGAGTGACCTGAAGAAGTGGTGACCGGAGCGCTTATGTGGAAACCAACCCGGTCC 434

RESULT 15
LOCUS    BY710826
DEFINITION  1012 bp mRNA linear EST 16-DEC-2002
CDNA clone 2610027010 5', mRNA sequence.
ACCESSION  BY710826
VERSION    BY710826.1 GI:27122059
KEYWORDS   EST.
SOURCE     Mus musculus (house mouse)
ORGANISM   Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 1012)
AUTHORS     Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S.,
            Nikaido, I., Osato, N., Saito, R., Suzuki, H., Yamanaka, I.,
            Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A.,
            Schonbach, C., Gojobori, T., Baldarelli, R., Hill, D. P., Bult, C.,
            Hume, D. A., Quackenbush, J., Schriml, L. M., Kanapin, A., Matsuda, H.,
            Batalov, S., Beisel, K. W., Blake, J. A., Brad, D., Bruscia, V.,
            Chothia, C., Corbani, L. E., Cousins, S., Dalla, E., Dragani, T. A.,
            Fietcher, C. F., Forrest, A., Frazer, K. S., Gaasterland, T.,
            Gariboldi, M., Gissi, C., Godzik, A., Gough, J., Grimmond, S.,
            Gustinch, S., Hirokawa, N., Jackson, I. J., Jarvis, E. D., Kanai, A.,
            Kawai, H., Kawasawa, Y., Kedzierski, R. M., King, B. L., Konagaya, A.,
            Kurochkin, I. V., Lee, Y., Lenhard, B., Lyons, P. A., Maglott, D. R.,
            Maltais, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T.,
            Numata, K., Okido, T., Pavan, W. J., Perte, G., Pesole, G.,
            Petrovsky, N., Pillai, R., Pontius, J. U., Qi, D., Ramachandran, S.,
            Ravasi, T., Reed, J. C., Reed, D. J., Reid, J., Ring, B. Z., Ringwald, M.,
            Sandelin, A., Schneider, C., Semple, C. A., Setou, M., Shimada, K.,
            Sultana, R., Takenaka, Y., Taylor, M. S., Teasdale, R. D., Tomita, M.,
            Verardo, R., Wagner, L., Wahlestedt, C., Wang, Y., Watanabe, Y.,
            Wells, C., Wilming, L. G., Wynshaw-Boris, A., Yanagisawa, M., Yang, I.,
            Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P.,
            Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura, M.,
            Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K.,
            Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Inetani, K., Ishii, Y.,
            Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K.,
            Shingawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E. S.,
            Rogers, J., Birney, B. and Hayashizaki, Y.
            Analysis of the mouse transcriptome based on functional annotation
            of 60,770 full-length cDNAs
            Nature 420, 563-573 (2002)
JOURNAL    2354583
MEDLINE    12466851
PUBMED
COMMENT    Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center (GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-resgsc.riken.jp, URL: http://genome.gsc.riken.jp/
            Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Carninci, P.,
            Fukuda, S., Hashizume, W., Hayashida, K., Hirozane, T., Hori, F.,
            Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kawai, J., Kojima, Y.,
            Kondo, S., Konno, H., Koya, S., Miyazaki, A., Murata, M., Nakamura, M.,
            Nemura, K., Numazaki, R., Ohno, M., Ohsato, N., Saito, R., Sakazume, N.,
            Sano, H., Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tagami, M.,

```

Takeda, Y., Waki, K., Watahiki, A., Muramatsu, M. and Hayashizaki, Y.  
 Direct Submission  
 Computational Analysis of Full-Length Mouse cDNAs Compared with  
 Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)  
 Normalization and subtraction of cap-trapper-selected cDNAs to  
 prepare full-length cDNA libraries for rapid discovery of new  
 genes. Genome Res. 10 (10), 1617-1630 (2000)  
 RIKEN integrated sequence analysis (RISA) system--384-format  
 sequencing pipeline with 384 multicapillary sequencer. Genome Res.  
 10 (11), 1757-1771 (2000)

Computer-based methods for the mouse full-length cDNA  
 encyclopedia: real-time sequence clustering for construction of a  
 nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
 cDNA library was prepared and sequenced in Mouse Genome  
 Encyclopedia Project of Genome Exploration Research Group in Riken  
 Genomic Sciences Center and Genome Science Laboratory in Riken.  
 Division of Experimental Animal Research in Riken contributed to  
 prepare mouse tissues.

Please visit our web site (<http://genome.gsc.riken.go.jp/>) for  
 further details.

## FEATURES

## Location/Qualifiers

```
1..1012
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="2610027010"
/sex="mixed"
/dev_stage="10 days embryo"
/lab_host="DH10B"
/clone_lib="RIKEN full-length enriched, 10 days embryo"
/notes="Site 1: SalI; Site 2: BamHI; cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN. Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer [5'
GAGAGAGAGAGATCCAGAGCTCTTTTCTTTTWTN 3'], cDNA was
prepared by using trehalose thermo-activated reverse
transcriptase and subsequently enriched for full-length by
cap-trapper. Second strand cDNA was prepared with the
primer adapter of sequence [5'
GAGAGAGATTCGAGTTAATAATATCCCCCCCCCC 3']. cDNA
was cloned into the XhoI and BamHI sites. Vector: a
modified pBluescript KS(+) after bulk excision from Lambda
FLC I"
```

## ORIGIN

```
Alignment Scores:
Pred. No.: 2,84e-83 Length: 1012
Score: 887.00 Matches: 172
Percent Similarity: 98.32% Conservative: 4
Best Local Similarity: 96.09% Mismatches: 3
Query Match: 96.73% Indels: 0
DB: 6 Gaps: 0
```

US-09-701-674A-23 (1-179) x BY170826 (1-1012)

```
QY 1 MetSerThrGlySerLeuSerAspValGluAspLeuGlnGluValGluMetLeuGluCys 20
Db 241 ATGTCACCTGGCTCCCTCAGCGATGTAGAAGACCTTCAGAGGTTGGAGATGCTGGACTGT 300
QY 21 AspGlyLeuLysMetAspSerAsnLysGluPheValThrSerAsnGluSerThrGluGlu 40
Db 301 GACTCCCTGAAGTGGACTCCACACAGAGAGTTTGGAACTTCCAACGAGACCCGAGGAG 360
QY 41 SerSerAsnCysGluAsnGlySerProGlnLysGlyArgGlyGlyLeuGlyLysArg 60
Db 361 GGCCTCCACATCGCAGAACGGGCTCTCCACAGAGGGTCCGGTGGCTAGCGAAGAGAGG 420
QY 61 LysAlaProThrLysLysSerProLeuSerGlyValSerGlnGluGlyLysGlnValGln 80
```

```
Db 421 AAGCGGCCCTACTAAGAAAAGCCCGCTCAGCGGGGTACGCCAGGAGGCGCAAGAGGTCCAG 480
QY 81 ArgAsnAlaAlaAsnAlaArgGluArgAlaArgMetArgValLeuSerLysAlaPheSer 100
Db 481 CGCAACGCGGCCCAATGCTCGTAGCGGGCCCGCATGCGGGGTGCTGAGCAAGGCCTTCTCC 540
QY 101 ArgLeuLysThrThrLeuProTyrValProProAspThrLysLysLeuSerLysLeuAspThr 120
Db 541 AGGCTCAAGACCCACCCCTGGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 600
QY 121 LeuArgLeuAlaSerSerTyrIleAlaHisLeuArgGlnIleLeuAlaAsnAspLysTyr 140
Db 601 CTCAGGCTGGCGTCCAGCTACATCGCTCACTTAAGGCAGATCCTGGCCCAACGACAAGTAC 660
QY 141 GluAsnGlyTyrIleHisProValAsnLeuThrTyrProPheMetValAlaGlyLysPro 160
Db 661 GAGAACCGTTTACATTCACCCAGTCAACCTGACGTGGGCCCTTTATGGTGGCGCGCAACCA 720
QY 161 GluSerAspLeuLysGluValValThrAlaSerArgLeuCysGlyThrThrAlaSer 179
Db 721 GAGATGACCTGAAGAGAGTGGTGACCGCCCAACCGCTTGTGTGGAATACAGCATCC 777
```

Search completed: December 17, 2004, 00:03:53  
 Job time : 2876 secs

This Page Blank (uspto)

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 16, 2004, 17:03:57 ; Search time 4403 Seconds  
(without alignments)

10527.218 Million cell updates/sec

Title: US-09-701-674A-54

Perfect score: 1272

Sequence: 1 tctacggccacgactctggg.....attgtataaaaaaaaaaaaaa 1272

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 32822875 seqs, 18219865908 residues

Total number of hits satisfying chosen parameters: 65645750

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

1: gb\_est1:\*

2: gb\_est2:\*

3: gb\_hic:\*

4: gb\_est3:\*

5: gb\_est4:\*

6: gb\_est5:\*

7: gb\_est6:\*

8: gb\_gss1:\*

9: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	1248	98.1	1252	3	CR616308 full-leng
2	1240	97.5	1250	3	CR606057 full-leng
3	1221	96.0	1221	3	CR594669 full-leng
C 4	1003.4	78.9	1085	1	AL574170 AL574170
5	995.8	78.3	1099	1	AL552119 AL552119
C 6	974	76.6	1024	1	AL375799 AL375799
7	963.8	75.8	1044	1	AL550251 AL550251
C 8	959.8	75.5	1169	1	AL552084 AL552084
9	939	73.8	965	1	AL552380 AL552380
10	931	73.2	1056	5	BM337699 BX337699
11	878.8	69.1	1038	4	BM544256 AGENCOURT
12	824	64.8	1098	4	BM923344 AGENCOURT
13	795	62.5	796	5	BM923344 AGENCOURT
14	788	61.9	996	5	BM923344 AGENCOURT
15	787	61.9	800	6	BM923344 AGENCOURT
16	770.4	60.6	794	6	CF127519 UI-HF-ETO
17	767.4	60.3	782	6	CF127034 UI-HF-ETO
18	763.6	60.0	1065	5	CF128484 UI-HF-ETO
19	752	59.1	818	4	BQ072411 AGENCOURT
20	751.2	59.1	1328	3	AK011575 Mus muscu
21	747.2	58.7	837	4	BI770596 603060402
22	747	58.7	826	4	BI768711 603057274
C 23	745.4	58.6	877	7	CK70542 958729 MA
24	738.8	58.1	757	6	CF126927 UI-HF-ETO

25 734 57.7 744 6 CF128914 UI-HF-ETO  
26 734 57.7 746 4 BI770998  
27 724.4 56.9 785 6 CF127896  
28 706.6 55.6 738 6 CF130905  
29 698.6 54.9 717 6 CF126467  
30 697.2 54.8 1167 4 BM811192  
31 696.8 54.8 718 7 CN409630  
32 683.8 53.8 852 7 CF994025  
33 681 53.5 696 5 BU677833  
34 673 52.9 677 5 BM975273  
35 670.4 52.7 779 4 BI837765  
36 668 52.5 815 2 AM025547  
37 664.4 52.2 771 2 AM242648  
38 663.6 52.1 784 7 CR770983  
39 652 51.3 851 5 BX337613  
40 648.4 51.0 653 5 BQ016290  
41 640 50.3 651 4 BI765033  
42 639 50.2 655 5 BQ019293  
43 629.2 49.5 694 4 BI838855  
44 621.8 48.9 827 5 BX928295  
45 621 48.8 924 4 BI763441

#### ALIGNMENTS

RESULT 1  
LOCUS CR616308 1252 bp mRNA linear HTC 21-JUL-2004  
DEFINITION full-length cDNA clone CSODI070VK16 of Placenta Cot 25-normalized of Homo sapiens (human).  
ACCESSION CR616308  
VERSION CR616308.1 GI:50497115  
KEYWORDS HTC; CNSUT\_cDNA.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
AUTHORS Li W.B., Gruber,C., Jessee,J. and Polayes,D.  
TITLE Full-length cDNA libraries and normalization  
JOURNAL Unpublished  
REMARK Contact : Feng Liang Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com/ Invitrogen Corporation 1600 Faraday Avenue  
REFERENCE 2 (bases 1 to 1252)  
AUTHORS Genoscope.  
JOURNAL Direct Submission  
SUBMITTED (20-JUL-2004) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqret@genoscope.cns.fr - Web : www.genoscope.cns.fr)  
COMMENT 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies, a division of Invitrogen.  
FEATURES  
Location/Qualifiers  
1..1252  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CSODI070VK16"  
/tissue\_type="Placenta Cot 25-normalized"  
/plasmid="pCMVSPORT\_6"

Query Match 98.1%; Score 1248; DB 3; Length 1252;  
Best Local Similarity 100.0%; Pred. No. 1.2e-287;  
Matches 1248; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 TCTACGCCACGACTCTGGAGTGGGAAACAGAGACCGGTTCTCTGTCGAGAGATC 60  
5 TCTACGCCACGACTCTGGAGTGGGAAACAGAGACCGGTTCTCTGTCGAGAGATC 64  
Db



```
QY 301 CTTCAAGAGTGGAGATGTTGGAAATGTGACGGGTGAAATGGATTGAAACAAGGAATTT 360
Db |||||
QY 311 CTTCAAGAGTGGAGATGTTGGAAATGTGACGGGTGAAATGGATTGAAACAAGGAATTT 370
Db |||||
QY 361 GTGACTTCCAAACGAGACACCGAGGAGCTCCAACCTGCGAGATGGTCTCCCGAGAG 420
Db |||||
QY 371 GTGACTTCCAAACGAGACACCGAGGAGCTCCAACCTGCGAGATGGTCTCCCGAGAG 430
Db |||||
QY 421 GCGCGCGCGCTGGCGCAAGAGAGGAGGAGCGGCCAACCAAGAGAGCGCCCTGAGCGGG 480
Db |||||
QY 431 GCGCGCGCGCTGGCGCAAGAGAGGAGGAGGAGCGGCCAACCAAGAGAGCGCCCTGAGCGGG 490
Db |||||
QY 481 GTGAGCAGAGGAGGAGAGCTCCAGCGCAACGCGGCCAACCGCGGAGAGCGGGCCCGC 540
Db |||||
QY 491 GTGAGCAGAGGAGGAGAGCTCCAGCGCAACGCGGCCAACCGCGGAGAGCGGGCCCGC 550
Db |||||
QY 541 ATGCGAGTGTGAGCAAGGCGCTTCTCCAGACTCAAGACCACTGCGCTGGGTGCGCCCGC 600
Db |||||
QY 551 ATGCGAGTGTGAGCAAGGCGCTTCTCCAGACTCAAGACCACTGCGCTGGGTGCGCCCGC 610
Db |||||
QY 601 GACACCAAGCTTCCAAAGCTGGACAGCTCAGGCTGCGCTGCGCTGCGCTGCGCTGCGCTG 660
Db |||||
QY 611 GACACCAAGCTTCCAAAGCTGGACAGCTCAGGCTGCGCTGCGCTGCGCTGCGCTGCGCTG 670
Db |||||
QY 661 AGCAGATCTGCTACGACCAATACGAGAACGGGTACATTACCCGGTCAACCTGACG 720
Db |||||
QY 671 AGCAGATCTGCTACGACCAATACGAGAACGGGTACATTACCCGGTCAACCTGACG 730
Db |||||
QY 721 TGGCCCTTTATGTTGGCGGGAACCCGAGAGTGACTGAAAGAGTGTGACCGGAGC 780
Db |||||
QY 731 TGGCCCTTTATGTTGGCGGGAACCCGAGAGTGACTGAAAGAGTGTGACCGGAGC 790
Db |||||
QY 781 CGCTTATGTGAAACCCCGGCTGCTGACCTGAGCTGAGGTGCGAGTCTGGAAAGCGCGCTC 840
Db |||||
QY 791 CGCTTATGTGAAACCCCGGCTGCTGACCTGAGCTGAGGTGCGAGTCTGGAAAGCGCGCTC 850
Db |||||
QY 841 CCGGGGGAGCGGGCGCCCGGGAAGGAGACCCCTGCGCTCAGTGTCTCTGTCTCTCTCTCT 900
Db |||||
QY 851 CCGGGGGAGCGGGCGCCCGGGAAGGAGACCCCTGCGCTCAGTGTCTCTGTCTCTCTCTCT 910
Db |||||
QY 901 CCCTCGCAATGCTCTCTCTCTGTCGACCCCGCGAGAACCTTTACACGACGAGGAG 960
Db |||||
QY 911 CCCTCGCAATGCTCTCTCTCTGTCGACCCCGCGAGAACCTTTACACGACGAGGAG 970
Db |||||
QY 961 ATTCTGTTCCAAACGAGAGATCAATTGCTACTTACAAAGATTTCCATCTATTAACTT 1020
Db |||||
QY 971 ATTCTGTTCCAAACGAGAGATCAATTGCTACTTACAAAGATTTCCATCTATTAACTT 1030
Db |||||
QY 1021 TATTAACTTCAACGATGACTCTGCAAGCCTTCTGCTCAAGTGAATATGTAAT 1080
Db |||||
QY 1031 TATTAACTTCAACGATGACTCTGCAAGCCTTCTGCTCAAGTGAATATGTAAT 1090
Db |||||
QY 1081 AATAATATAAATAGATAGGCTTCAATGATCTTTTGTACAAATATGTTAAAT 1140
Db |||||
QY 1091 AATAATATAAATAGATAGGCTTCAATGATCTTTTGTACAAATATGTTAAAT 1150
Db |||||
QY 1141 GTAGATCATAGGATAGTCACTTTGACAGTCAATTTATAAGTAAATTCACCTTAAAGATA 1200
Db |||||
QY 1151 GTAGATCATAGGATAGTCACTTTGACAGTCAATTTATAAGTAAATTCACCTTAAAGATA 1210
Db |||||
QY 1201 TATATTTTTCACCAAGTTTGTCTACTTTTGAAATAA 1240
Db |||||
QY 1211 TATATTTTTCACCAAGTTTGTCTACTTTTGAAATAA 1250
Db |||||
```

```
RESULT 3
CR594669
LOCUS CR594669
DEFINITION full-length cDNA clone CS0D1059YH17 of Placenta Cot 25-normalized
of Homo sapiens (human).
ACCESSION CR594669
VERSION CR594669.1 GI:50475476
KEYWORDS HTC; CNSLT_cDNA.
```

```
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 1221)
AUTHORS Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished
REMARK Contact : Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Paraday Avenue
REFERENCE 2 (bases 1 to 1221)
GENOSCOPE Genoscope.
DIRECT SUBMISSION Direct Submission
JOURNAL Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqrefgenoscope.cns.fr)
- Web : www.genoscope.cns.fr)
COMMENT 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
ended enriched, double-strand cDNA was digested with Not I and cloned
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
was normalized. Library was constructed by Life Technologies, a
division of Invitrogen.
FEATURES
Location/Qualifiers
1..1221
source
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0D1059YH17"
/tissue_type="Placenta Cot 25-normalized"
/plasmid="pCMVSPORT_6"
ORIGIN
Query Match 96.0%; Score 1221; DB 3; Length 1221;
Best Local Similarity 100.0%; Pred.No.3.6e-281;
Matches 1221; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 9 CACGACTCTGGAGTGGGAAACAGAGAGCGGTTCTCTCTCAGAGGCTTCGGTTTCAGGCTCTCT 68
Db 1 CACGACTCTGGAGTGGGAAACAGAGAGCGGTTCTCTCTCAGAGGCTTCGGTTTCAGGCTCTCT 60
QY 69 TCCTTCTCACAACCTCTCGAAGGGGAAAGGTTGTGAGACCCCAACAGAGACCCCAACTCCA 128
Db 61 TCCTTCTCACAACCTCTCGAAGGGGAAAGGTTGTGAGACCCCAACAGAGACCCCAACTCCA 120
QY 129 GTCCTCCAGAGAGGTGGTCTGGCCACACTCGGAGGCGCTTCGGTTTCAGGCTCTCTCT 188
Db 121 GTCCTCCAGAGAGGTGGTCTGGCCACACTCGGAGGCGCTTCGGTTTCAGGCTCTCTCT 180
QY 189 GTCTCTCTCTCACCTCTTCTCTGCTTTCTCTGCTCTCTCTCTCTCTCTCTCTCTCTCT 248
Db 181 GTCTCTCTCTCACCTCTTCTCTGCTTTCTCTGCTCTCTCTCTCTCTCTCTCTCTCTCT 240
QY 249 CTGCTCACTCTCCCAACATGTCCACGGCTCCCTCAGCGATGTGGAGACCTTCAAGA 308
Db 241 CTGCTCACTCTCCCAACATGTCCACGGCTCCCTCAGCGATGTGGAGACCTTCAAGA 300
QY 309 GGTGAGATGTTGGAATGTGACGGGTTGAAATCGAATCGAATCGAATTTGACTTC 368
Db 301 GGTGAGATGTTGGAATGTGACGGGTTGAAATCGAATCGAATTTGACTTC 360
QY 369 CAACGAGAGCACCAGAGAGAGTCCAACTCGAGAATGGGTCTCCCGAGAGGCGCGG 428
Db 361 CAACGAGAGCACCAGAGAGAGTCCAACTCGAGAATGGGTCTCCCGAGAGGCGCGG 420
QY 429 CGGCTGGGCAAGAGAGGAGGCGCCACACAGAGAGCCCTGAGCGGGTCAAGCA 488
Db 421 CGGCTGGGCAAGAGAGGAGGCGCCACACAGAGAGCCCTGAGCGGGTCAAGCA 480
QY 489 GGAGGGGAAGCAGGTCCAGCGCAACGCGCGCAACGCGGAGAGCGGCGCGCATGCGAGT 548
Db 481 GGAGGGGAAGCAGGTCCAGCGCAACGCGCGCAACGCGGAGAGCGGCGCGCATGCGAGT 540
QY 549 GCTGAGCAAGGCTTCTCCAGACTCAAGACCACTGCCCTGGGTGGTGGCCCGCACCA 608
```









source	1.1044	/organism="Homo sapiens"	
		/mol_type="mRNA"	
		/db_xref="taxon:9606"	
		/clone="CS0D1039YJ05"	
		/tissue_type="PLACENTA COT 25-NORMALIZED"	
		/clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"	
		/note="1st strand cDNA was primed with a NotI-oligo (dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."	
ORIGIN			
Query Match	75.8%;	Score 963.8;	DB 1; Length 1044;
Best Local Similarity	97.9%;	Pred. No. 1.3e-219;	
Matches 1000;	Conservative	7; Mismatches	9; Indels 5; Gaps 3;
QY	1	TCTACGGCCACGACTCTGGAGTGGGAAACAGAGACCGGTTCTCTCTGCTCAGAGTC	60
Db	11	TCTACGGCCACGACTCTGGAGTGGGAAACAGAGACCGGTTCTCTCTGCTCAGAGTC	70
QY	61	CTCGGGTTCTTCTCACAACTCTGCGAAGGGGAAAGGTTGTGAGACCAACAGACCC	120
Db	71	CTCGGGTTCTTCTCACAACTCTGCGAAGGGGAAAGGTTGTGAGACCAACAGACCC	130
QY	121	CAACTCAGCTCCACAGAGAGTGGCTGGCCACACTCGGAGGGCTTTGGTTTCAGG	180
Db	131	CAACTCAGCTCCACAGAGAGTGGCTGGCCACACTCGGAGGGCTTTGGTTTCAGG	190
QY	181	GTCTCTGCTCTCTCTCACCTCTTCTCTGCTTCTCTCTCTCTCTCTCTCTCT	240
Db	191	GTCTCTGCTCTCTCTCACCTCTTCTCTGCTTCTCTCTCTCTCTCTCTCTCT	250
QY	241	CT	300
Db	251	CT	310
QY	301	CTTCAAGAGTGAGATGTGGAATGTGACGGTGTGAATGTGGAATGTGGAATGTG	360
Db	311	CTTCAAGAGTGAGATGTGGAATGTGACGGTGTGAATGTGGAATGTGGAATGTG	370
QY	361	GTGACTTCCACAGAGACCGAGGAGCTTCAACTGGGAGATGGGTCTCCCAAGAG	420
Db	371	GTGACTTCCACAGAGACCGAGGAGCTTCAACTGGGAGATGGGTCTCCCAAGAG	430
QY	421	GGCGGGGGCTGGCGAAGAGGAGAGGCGCCCAACAGAGAGCCCTCTGAGCGG	480
Db	431	GGCGGGGGCTGGCGAAGAGGAGAGGCGCCCAACAGAGAGCCCTCTGAGCGG	490
QY	481	GTGAGCCAGAGGGGAGAGGTCCAGCGCAACGCGCCCAACGCGAGAGCGGCGCGC	540
Db	491	GTGAGCCAGAGGGGAGAGGTCCAGCGCAACGCGCCCAACGCGAGAGCGGCGCGC	550
QY	541	ATGCGAGTGTGAGGAAGGCTTCTCCAGCTCAAGACACACCGCTGGCTGGTGGCTCC	600
Db	551	ATGCGAGTGTGAGGAAGGCTTCTCCAGCTCAAGACACACCGCTGGCTGGTGGCTCC	610
QY	601	GACACCAAGCTCTCCAGGTGACAGCTCAGCTGGCTTCCAGCTTACATGCCCACTTG	660
Db	611	GACACCAAGCTCTCCAGGTGACAGCTCAGCTGGCTTCCAGCTTACATGCCCACTTG	670
QY	661	AGCAGATCTTGCTTAACGAAATACGAGAGCGGTATCATTCACCGGTCAACCTGACG	720
Db	671	AGCAGATCTTGCTTAACGAAATACGAGAGCGGTATCATTCACCGGTCAACCTGACG	730
QY	721	TGCGCTTTATGTGCGCGGAAACCGAGAGTGACCTGAAGAGTGTGTGACCGGAGC	780
Db	731	TGCGCTTTATGTGCGCGGAAACCGAGAGTGACCTGAAGAGTGTGTGACCGGAGC	790
QY	781	CGTTATGTGGAACCAACCGCGTCTCTGACCTTGGAGTGGAGTCTGGGAAAGCGCGTC	840
Db	791	CGTTATGTGGAACCAACCGCGTCTCTGACCTTGGAGTGGAGTCTGGGAAAGCGCGTC	850
QY	841	CCGGGGGAGCGGGCCCGGGAGCGACCCCTGCTCAGTGTCTCTCTCTCTCTCTCT	900
Db	851	CCGGGGGAGCGGGCCCGGGAGCGACCCCTGCTCAGTGTCTCTCTCTCTCTCTCT	910
QY	901	CCCTCGCAATGCT	960
Db	911	CCCTCGCAATGCT	969
QY	961	ATTGCTTTCCAAACACAGAGAGATCAATTGTACTTACAAAGATTCCCATCTATTAACTT	1020
Db	970	ATTGCTTTCCAAACACAGAGAGATCAATTGTACTTACAAAGATTCCCATCTATTAACTT	1025
QY	1021	T 1021	
Db	1026	W 1026	
RESULT 8			
AL552084/c			
LOCUS	AL552084	Homo sapiens	PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
DEFINITION	clone CS0D1059YH17 3-PRIME, mRNA sequence.		
ACCESSION	AL552084		
VERSION	AL552084.3	GI:45856873	
KEYWORDS	EST.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
AUTHORS	1. (Bases 1 to 1169)		
TITLE	Li, W. B., Gruber, C., Jessee, J. and Polayes, D.		
JOURNAL	Full-length cDNA libraries and normalization		
COMMENT	Unpublished (2001)		
	On Feb 15, 2001 this sequence version replaced gi:31273900.		
	Contact: Genoscope		
	Genoscope - Centre National de Sequencage		
	BP 191 91006 EVRY cedex - France		
	Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr		
	1st strand cDNA was primed with a NotI-oligo (dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies, a division of Invitrogen. This sequence belongs to sequence cluster 3812.f		
	For more information about this cluster, see		
	http://www.genoscope.cns.fr/cdna?s=CS0D1059CD09NFl&e=3812.f.		
FEATURES	Location/Qualifiers		
source	1..1169		
	/organism="Homo sapiens"		
	/mol_type="mRNA"		
	/db_xref="taxon:9606"		
	/clone="CS0D1059YH17"		
	/tissue_type="PLACENTA COT 25-NORMALIZED"		
	/clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"		
	/note="1st strand cDNA was primed with a NotI-oligo (dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."		
ORIGIN			
Query Match	75.5%;	Score 959.8;	DB 1; Length 1169;
Best Local Similarity	97.3%;	Pred. No. 1.2e-218;	
Matches 988;	Conservative	4; Mismatches	21; Indels 2; Gaps 2;
QY	207	TCCTCGGTTTCT	266
Db	1019	TCCTCGGTTTCT	961
QY	267	CATGTCCACCGGCTCCCTCAGCGATGTGGAGACCTTCAAGAGTGGAGATGTTGGAATG	326
Db	960	CATGTCCACCGGCTCCCTCAGCGATGTGGAGACCTTCAAGAGTGGAGATGTTGGAATG	901
QY	327	TGACGGGTTGAAATGGATTGGAACAGGAATTTGTGACTTCCACAGAGACCCGAGGA	385



QY 601 GACACCAAGCTCTCCAAAGCTGGACACGCTCAGGCTGGCGTCCAGCTACATCGGCCACTTG 660  
 Db 605 GACACCAAGCTCTCCAAAGCTGGACACGCTCAGGCTGGCGTCCAGCTACATCGGCCACTTG 664  
 QY 661 AGGCAGATCTGGCTACGACCAATACGAAACGGGTACATTCACCGGTCAACCTGACG 720  
 Db 665 AGGCAGATCTGGCTACGACCAATACGAAACGGGTACATTCACCGGTCAACCTGACG 724  
 QY 721 TGGCCCTTTATGTGGCGGGAACCCGAGAGTCACTGAAAGAGTGGTGAACCGGAGC 780  
 Db 725 TGGCCCTTTATGTGGCGGGAACCCGAGAGTCACTGAAAGAGTGGTGAACCGGAGC 783  
 QY 781 CGCTTATGTGAACACCGGCTCTGACCTTGAGAGTGGAGTCTGGGAAAGCGCGCTC 840  
 Db 784 CGCTTATGTGAACACCGGCTCTGACCTTGAGAGTGGAGTCTGGGAAAGCGCGCTC 842  
 QY 841 CCGGGGGAGCGGGCCCGGGGAAGGACCCCTGCCCTCAGTCTCTCTGTCTCTCTCTC 900  
 Db 843 CCGGGGGAGCGGGCCCGGGGAAGGACCCCTGCCCTCAGTCTCTCTGTCTCTCTCTC 902  
 QY 901 CCCTCGCAATGCT 960  
 Db 903 CCCTCGCAATGCT 962  
 QY 961 ATT 963  
 Db 963 ATT 965

RESULT 10  
 BX337699  
 LOCUS  
 DEFINITION BX337699 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA  
 clone CSODI048YE15 5'-PRIME, mRNA sequence.  
 ACCESSION BX337699  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM Homo sapiens (human)  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 1056)  
 Li, W.B., Gruber, C., Jessee, J., and Polayes, D.  
 Full-length cDNA libraries and normalization  
 Unpublished (2001)  
 On May 2, 2003 this sequence version replaced gi:30345636.  
 Contact: Genoscope  
 Genoscope - Centre National de Sequencage  
 BP 191 91006 EVRY cedex - France  
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr  
 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime  
 end enriched, double-strand cDNA was digested with Not I and cloned  
 into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library  
 was normalized. Library was constructed by life technologies, a  
 division of Invitrogen. This sequence belongs to sequence cluster  
 3812.f  
 For more information about this cluster, see  
 http://www.genoscope.cns.fr/cdna?s=CSODI048AC08QP1&c=3812.f.  
 Location/Qualifiers

FEATURES  
 source  
 1..1056  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="CSODI048YE15"  
 /tissue\_type="PLACENTA COT 25-NORMALIZED"  
 /clone\_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"  
 /note="1st strand cDNA was primed with a NotI-oligo(dT)  
 primer. Five prime end enriched, double-strand cDNA was  
 digested with Not I and cloned into the Not I and EcoR V  
 sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN

Query Match 73.2%; Score 931; DB 5; Length 1056;  
 Best Local Similarity 96.9%; Pred. No. 9.2e-212;  
 Matches 953; Conservative 14; Mismatches 14; Indels 2; Gaps 2;  
 QY 1 TCTAGGGCCACGACTCTGGAGTGGGAAACAGAGAGCCGGTTCCTCTGCTGCAGAAATC 60  
 Db 11 TCTAGGGCCACGACTCTGGAGTGGGAAACAGAGAGCCGGTTCCTCTGCTGCAGAAATC 70  
 QY 61 CTCGGGGTTCCCTTCTCACAACCTCTCGAAGGGGAAAGGTTGTGAGACCCCAACAGAGCC 120  
 Db 71 CTCGGGGTTCCCTTCTCACAACCTCTCGAAGGGGAAAGGTTGTGAGACCCCAACAGAGCC 130  
 QY 121 CAACTCCAGCTCCCAAGAGAGAGTGGTGCACCACTCTCGGAGAGGCTCTTGGTTTCAAG 180  
 Db 131 CAACTCCAGCTCCCAAGAGAGAGTGGTGCACCACTCTCGGAGAGGCTCTTGGTTTCAAG 190  
 QY 181 GTCTCTGTCT 240  
 Db 191 GTCTCTGTCT 250  
 QY 241 CT 300  
 Db 251 CT 310  
 QY 301 CTTCAGAGGTGGAGATGTTGGAATGTGACGGGTTGAAATGGAATGGAATGGAATGGA 360  
 Db 311 CTTCAGAGGTGGAGATGTTGGAATGTGACGGGTTGAAATGGAATGGAATGGAATGGA 370  
 QY 361 GTGACTTCCAAACGAGAGCACCAGAGAGAGTCCAACTCCGAGAAATGGGTCTCTCCCAAG 420  
 Db 371 KTGACTTCCAAACGAGAGCACCAGAGAGAGTCCAACTCCGAGAAATGGGTCTCTCCCAAG 430  
 QY 421 GCGCGCGCGGCTGGGCAAG 480  
 Db 431 GCGCGCGCGGCTGGGCAAG 490  
 QY 481 GTGAGCCAGAGAGGAGAGAGAGTCCAGCGCAACCGCGCAACCGCGCAACCGCGCAAC 540  
 Db 491 GTGAGCCAGAGAGGAGAGAGTCCAGCGCAACCGCGCAACCGCGCAACCGCGCAAC 550  
 QY 541 ATGGAGTGTGAGAGAGGCTTCTCAGACTCAAGACCAACCTGCGCTGGGTGCCCGCC 600  
 Db 551 ATGGAGTGTGAGAGAGGCTTCTCAGACTCAAGACCAACCTGCGCTGGGTGCCCGCC 610  
 QY 601 GACACCAAGCTCTCCAAAGTGGACACGCTCAGGCTGGGCTCCAGCTACATCGCCCACTTG 660  
 Db 611 GACACCAAGCTCTCCAAAGTGGACACGCTCAGGCTGGGCTCCAGCTACATCGCCCACTTG 670  
 QY 661 AGGCAGATCTCTGGCTAACGACAAATACGAGAACGGGTACATTCACCGGTCAACCTGACG 720  
 Db 671 AGGCAGATCTCTGGCTAACGACAAATACGAGAACGGGTACATTCACCGGTCAACCTGACG 730  
 QY 721 TGGCCCTTTATGTGGCGGGAACCCGAGAGTCACTGAAAGAGTGGTGAACCGGAGC 780  
 Db 731 TGGCCCTTTATGTGGCGGGAACCCGAGAGTCACTGAAAGAGTGGTGAACCGGAGC 790  
 QY 781 CGCTTATGTGAACACCGGCT 840  
 Db 791 CGCTTATGTGAACACCGGCT 850  
 QY 841 CCGGGGGAGCGGGCCCGGGGAAGGACCCCTGCCCTCAGTCTCTCTCTCTCTCTCTCTCT 900  
 Db 851 CCGGGGGAGCGGGCCCGGGGAAGGACCCCTGCCCTCAGTCTCTCTCTCTCTCTCTCTCT 909  
 QY 901 CCCCTCGCAATGCT 960  
 Db 910 CCCCTCGCAATGCT 968  
 QY 961 ATTGCTTCCAAACGAGAGGAGA 983  
 Db 969 WTCGTTTCCAAACARAGGGGNTA 991





/db\_xref="taxon:9606"  
/clone="IMAGE:5758942"  
/lab\_host="DH10B"  
/clone\_lib="NIH\_MGC\_116"  
/note="Organ: pocco colon, kidney, stomach; Vector: PCWV-SPORE; Site 1: NotI; Site 2: EcoRV (destroyed); RNA source anonymous pool of 3 colons, age 26 yo male, 49 yo female, 71 yo male colon; 46 yo male kidney, and pool of 2 stomachs, 62 yo male and 70 yo female. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.4 kb, insert size range 1-3 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 023. Note: this is a NIH\_MGC Library."

## ORIGIN

Query Match 64.8%; Score 824; DB 4; Length 1098;  
Best Local Similarity 95.1%; Pred. No. 3.8e-186;  
Matches 84; Conservative 0; Mismatches 15; Indels 1; Gaps 1;  
QY 14 CTCTGGAGTGGGAAACAGAGCCGGTTCCTCTGTCGAGAGTCTCGGGTTCCTT 73  
Db 14 CTCTGGAGTGGGAAACAGAGCCGGTTCCTCTGTCGAGAGTCTCGGGTTCCTT 73  
QY 74 CTCACACTCTGCGAAGGGAAAGGGTTGTGACCCCAACAGACCCCAACTCCAGCTCC 133  
Db 74 CTCACACTCTGCGAAGGGAAAGGGTTGTGACCCCAACAGACCCCAACTCCAGCTCC 133  
QY 134 CAGCAGAGGTGGTGGCCACACTCGGAGGCTCTTGGTTTCAGGGTCTCTCTCTC 193  
Db 134 CAGCAGAGGTGGTGGCCACACTCGGAGGCTCTTGGTTTCAGGGTCTCTCTCTC 193  
QY 194 TCTCTACCTCTTCT 253  
Db 194 TCTCTACCTCTTCT 253  
QY 254 CCATCCCCCAACATGTCACCGTCTCCCTAGCAGTGTGAGGACCTTCAGAGGTGG 313  
Db 254 CCATCCCCCAACATGTCACCGTCTCCCTAGCAGTGTGAGGACCTTCAGAGGTGG 313  
QY 314 AGATGTTGGAATGTGAGCGGTGAAATGATTCGAACAGGAATTTGTACTTCCAGC 373  
Db 314 AGATGTTGGAATGTGAGCGGTGAAATGATTCGAACAGGAATTTGTACTTCCAGC 373  
QY 374 AGAGCCAGGAGAGAGCTCAATCGCAGAAATGGGTCTCCCAAGAGGCGCGCGGCC 433  
Db 374 AGAGCCAGGAGAGAGCTCAATCGCAGAAATGGGTCTCCCAAGAGGCGCGCGGCC 433  
QY 434 TGGGCAAGAGAGGAGGCGCCCAACAGAGAGCCCTGAGCGGGTCAGCCAGGAGG 493  
Db 434 TGGGCAAGAGAGGAGGCGCCCAACAGAGAGCCCTGAGCGGGTCAGCCAGGAGG 493  
QY 494 GGAAGCAGGTTCAGCGCAACGCGCCCAACGCGAGAGCGGGCCCGCATGCGAGTCTGA 553  
Db 494 GGAAGCAGGTTCAGCGCAACGCGCCCAACGCGAGAGCGGGCCCGCATGCGAGTCTGA 553  
QY 554 GCAAGCGCTTCTCCAGACTCAGACCACTGCGCTGGGTGGCGCCCGCCAGCAAGCTCT 613  
Db 554 GCAAGCGCTTCTCCAGACTCAGACCACTGCGCTGGGTGGCGCCCGCCAGCAAGCTCT 613  
QY 614 CCAAGCTGAGCAGCTCAGGCTGGCGTCCAGGTATCATCGCCCACTTGAAGCAGATCCTGG 673  
Db 614 CCAAGCTGAGCAGCTCAGGCTGGCGTCCAGGTATCATCGCCCACTTGAAGCAGATCCTGG 673  
QY 674 CTACACAAATACGAGAACGGGTACATTCACCGGTCAACTGAGCTGG-CCCTTATG 732  
Db 674 CTACACAAATACGAGAACGGGTACATTCACCGGTCAACTGAGCTGG-CCCTTATG 732  
QY 733 GTGGCGGGAACCCGAGAGTGACCTGAAAGAGGTGGTCAACCGCAGCGGCTTATGTGA 792  
Db 734 GTGGCGGGAACCCGAGAGTGACCTGAAAGAGGTGGTCAACCGCAGCGGCTTATGTGA 793

QY 793 ACCACCGGCTCTGACCTTGGAGTGGAGTCTGGAAAGGCGGCTCCCGGGGGAGCG 852  
Db 794 ACCACCGGCTCTGACCTTGGAGTGGAGTCTGGAAAGGCGGCTCTCCCGGGGGAGCG 853  
QY 853 GGCCCCCGGGAAGCGACCCC 872  
Db 854 GGCCCCCGGGAAGCGCACCC 873  
EX107489 796 bp mRNA linear EST 06-FEB-2003  
BX107489 Soares placenta Nb2HP Homo sapiens cDNA clone  
IMAGE998N24234 ; IMAGE:151511, mRNA sequence.  
EX107489  
EX107489.1 GI:27834707  
EST.  
Homo sapiens (human)  
SOURCE Homo sapiens  
ORGANISM  
KEYWORDS  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 796)  
REFERENCE Ebert, L., Heil, O., Hennig, S., Neubert, P., Partsch, E., Peters, M.,  
Radelof, U., Schneider, D. and Korn, B.  
Human Unigeneset - RZPD3  
Unpublished (2003)  
Contact: Ina Rofls  
RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH  
Im Neuenheimer Feld 580, D-69120 Heidelberg, Germany  
RZPD; IMAGE998N24234.  
RZPDLIB; I.M.A.G.E. cDNA Clone Collection;  
Human Unigeneset - RZPD3 (RZPDLIB No.972)  
http://www.rzpd.de/CloneCards/cgi-  
bin/showlib.pl.cgi/response?libNo=972 Contact: Ina Rofls  
RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH  
Heubnerweg 6, D-14059 Berlin, Germany  
Tel: +49 30 32639 101  
Fax: +49 30 32639 111  
www.rzpd.de  
This clone is available royalty-free from RZPD;  
contact RZPD (clone@rzpd.de) for further information. Seq primer:  
M13r, Primer sequence: TTTACACAGGAACAGCTATGAC.

FEATURES  
source

Location/Qualifiers  
1..796  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE998N24234 ; IMAGE:151511"  
/sex="Female"  
/dev\_stage="placenta obtained at birth (full term)"  
/lab\_host="DH10B (ampicillin resistant)"  
/clone\_lib="Soares placenta Nb2HP"  
/note="Organ: placenta; Vector: pT7T3D (Pharmacia) with a  
modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st  
strand cDNA was primed with a Not I - oligo(dT) primer [5'  
AATCGAAGATTCGCGGCGGCGAGAAATTTTCTTTTCTTTT 3']  
double-stranded cDNA was ligated to Eco RI adaptors  
(Pharmacia), digested with Not I and cloned into the Not I  
and Eco RI sites of the modified pT7T3 vector. Library  
went through one round of normalization. Library  
constructed by Bento Soares and M.Fatima Bonaldo."

## ORIGIN

Query Match 62.5%; Score 795; DB 5; Length 796;  
Best Local Similarity 100.0%; Pred. No. 3.2e-179;  
Matches 795; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 74 CTCACACTCTGCGAAGGGAAAGGGTTGTGAGACCCCAACAGACCCCAACTCCAGTCC 133  
Db 1 CTCACACTCTGCGAAGGGAAAGGGTTGTGAGACCCCAACAGACCCCAACTCCAGTCC 60  
QY 134 CAGCAGAGGTGGCTGGCCACACTCGGAGGCTCTTGGTTTCAGGGTCTCTCTGCTCTC 193





QY 736 GCCGGAAACCCGAGAGTGACCTGAAGAAGTGGTGAACCGGAGCGGCTTATGTGAACC 795  
 Db 721 GCCGGAAACCCGAGAGTGACCTGAAGAAGTGGTGAACCGGAGCGGCTTATGTGAAC-C 779  
 QY 796 ACCGCTCTCTGACCTTGGAGTGGCAGTCTGGGAAGGCGCTCCCG 843  
 Db 780 ACCGCTCTCTGACCTTGGAGTGGTACCTGGAGGAGCTTCTCCG 827

## RESULT 15

CF127519 800 bp mRNA linear EST 05-AUG-2003  
 UI-HF-ET0-awh-f-14-0-UI-r1 NIH\_MGC\_214 Homo sapiens cDNA clone  
 IMAGE:30554533 5', mRNA sequence.

ACCESSION CF127519  
 VERSION CF127519.1 GI:33205841  
 KEYWORDS EST.

SOURCE Homo sapiens (human)

## ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 Bonaldo,M.F., Lennon,G. and Soares,M.B.  
 Normalization and subtraction: two approaches to facilitate gene  
 discovery

JOURNAL Genome Res. 6 (9), 791-806 (1996)

MEDLINE 97044477

PUBMED 889548

## COMMENT

Contact: Soares, MB  
 Coordinated Laboratory for Computational Genomics  
 University of Iowa  
 375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA  
 Tel: 319 335 8250  
 Fax: 319 335 9565

Email: bento-soares@uiowa.edu

Tissue Procurement: Mary Hendrix

cDNA Library prepared by: Dr. M. Bento Soares, University of Iowa  
 cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
 Clone Distribution: Distribution information can be found at  
<http://genome.uiowa.edu/distribution/humanfl.html>

The following repetitive elements were found in this cDNA  
 sequence: 172-253, >(GA)n#simple\_repeat (matched complement)

Seq primer: pYX-5.

## FEATURES

Source

1..800  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:30554533"  
 /tissue\_type="Chondrosarcoma Lung Metastasis cell lines"  
 /lab\_host="PH103 (T1 phage resistant)"  
 /clone\_lib="NIH MGC\_214"  
 /note="Organ: Lung; Vector: pYX-Asc; Site:1: EcoR I;  
 Site:2: Not I; The library was constructed according  
 Bonaldo, Lennon and Soares, Genome Research, 6:791-806,  
 1996. Denatured RNA was size fractionated on a 1% agarose  
 gel. First strand cDNA synthesis was primed with oligo-dT  
 primer containing a Not I site. Double strand cDNA was  
 size selected according to mRNA size fraction, ligated  
 with EcoR I adaptor, digested with Not I and then cloned  
 directionally into pYX-Asc vector. The library tag  
 sequence located between the Not I site and the polyA tail  
 is GATAGGCCA. Tissue was provided by Mary Hendrix."

## ORIGIN

Query Match 61.9%; Score 787; DB 6; Length 800;  
 Best Local Similarity 100.0%; Pred. No. 2.6e-177;  
 Matches 787; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 11 CGACTCTGGAGTGGGAAACACAGAGCGCGTCTCTCTGCTGCAGAAAGTCTCGGGGTTTC 70  
 Db 1 CGACTCTGGAGTGGGAAACACAGAGCGCGTCTCTCTGCTGCAGAAAGTCTCTCGGGGTTTC 60

Search completed: December 16, 2004, 20:16:36  
 Job time : 4410 secs

QY 71 CTTCTCAAACTCTCGAAGGGGAAAGGGTTGTGAGACCCAAACAGACCCCAATCCAGC 130  
 Db 61 CTTCTCAAACTCTCGAAGGGGAAAGGGTTGTGAGACCCCAACAGACCCCAATCCAGC 120  
 QY 131 TCCAGAGAGAGGTGGCTGCGGCACACTCGGAGAGGCTCTTGGTTTCAGGGTCTCTGT 190  
 Db 121 TCCAGAGAGAGGTGGCTGCGGCACACTCGGAGAGGCTCTTGGTTTCAGGGTCTCTGT 180  
 QY 191 CTCTCTCTCACCCCTCTTCTCTCGCTTTCTCTGTCTCTCTCTCTCTCTCTCTCTCT 250  
 Db 181 CTCTCTCTCACCCCTCTTCTCTCGCTTTCTCTGTCTCTCTCTCTCTCTCTCTCT 240  
 QY 251 CGTCCACTTCCCCCAAAACATGTCACCGGCTCCCTCAGCGATGTGAGAGACCTTCAAGAG 310  
 Db 241 CGTCCACTTCCCCCAAAACATGTCACCGGCTCCCTCAGCGATGTGAGAGACCTTCAAGAG 300  
 QY 311 TGGAGATGTTGGAATGTGACGGGTGAAATGATTCGAACAAGGAATTTGTGACTTCCA 370  
 Db 301 TGGAGATGTTGGAATGTGACGGGTGAAATGATTCGAACAAGGAATTTGTGACTTCCA 360  
 QY 371 ACAGAGCACCCGAGGAGAGCTCCAACCTGCGAGAAATGGGTCTCCCAAGAGGGCGGGCG 430  
 Db 361 ACAGAGCACCCGAGGAGAGCTCCAACCTGCGAGAAATGGGTCTCCCAAGAGGGCGGGCG 420  
 QY 431 GCCTGGCAGAGAGAGAGAGCGGCCACCAAGAGAGCCCTGAGGGGGTTCAGCCAGG 490  
 Db 421 GCCTGGCAGAGAGAGAGAGCGGCCACCAAGAGAGCCCTGAGGGGGTTCAGCCAGG 480  
 QY 491 AGGGGAAGCAGGTCCAGCGCAACCGCGCCCAACCGCGGAGAGCGGGCCCGCATCGAGTGC 550  
 Db 481 AGGGGAAGCAGGTCCAGCGCAACCGCGCCCAACCGCGGAGAGCGGGCCCGCATCGAGTGC 540  
 QY 551 TGAGCAAGGCTTCTCCAGACTCAAGACACACCTGCGCTGGTGGTGGTGGTGGTGGTGG 610  
 Db 541 TGAGCAAGGCTTCTCCAGACTCAAGACACACCTGCGCTGGTGGTGGTGGTGGTGGTGG 600  
 QY 611 TCTCCAAGCTGGACACGCTCAGGCTGGCGTCCAGTACATCGCCCACTTGAGGACAGATCC 670  
 Db 601 TCTCCAAGCTGGACACGCTCAGGCTGGCGTCCAGTACATCGCCCACTTGAGGACAGATCC 660  
 QY 671 TGGCTAACGACAAATACGAGACGGGTACATTACCCGGTCAACCTGACCTGACCTGACCT 730  
 Db 661 TGGCTAACGACAAATACGAGACGGGTACATTACCCGGTCAACCTGACCTGACCTGACCT 720  
 QY 731 TGGTGGCGGGGAAACCCGAGAGTGACCTGAAAGAGTGGTGAACCGGAGCGCGCTTATGTG 790  
 Db 721 TGGTGGCGGGGAAACCCGAGAGTGACCTGAAAGAGTGGTGAACCGGAGCGCGCTTATGTG 780  
 QY 791 GAACCCAC 797  
 Db 781 GAACCCAC 787

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: December 16, 2004, 21:52:41 ; Search time 88 Seconds

(without alignments)  
1445.809 Million cell updates/sec

Title: US-09-701-674A-23

Perfect score: 917

Sequence: 1 MSGSLSDVEDQEVEMLEC.....PESDLKEVTSRLCGTTAS 179

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5

Ygapop 10.0 , Ygapext 0.5

Fgapop 6.0 , Fgapext 7.0

Delop 6.0 , Delext 7.0

Searched: 824507 seqs, 355394441 residues

Total number of hits satisfying chosen parameters: 1649014

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODE=frame+p2n.model -DRV=xlh  
-Q/cgn2\_1/USPTO.spool\_h/US09701674/runat\_15122004\_100441\_27529/app\_query.fasta\_1.327  
-DB=Issued Patents NA -QFWT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPEL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR SCORE=ptc -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTPMT=ptc -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09701674 @CNC 1.1 69 @runat\_15122004\_100441\_27529 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSFBLOCK=100 -LONGLOG  
-DSV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents NA:\*

1: /cgn2\_6/ptodata/1/ina/5A\_COMB.seq:\*  
2: /cgn2\_6/ptodata/1/ina/5B\_COMB.seq:\*  
3: /cgn2\_6/ptodata/1/ina/6A\_COMB.seq:\*  
4: /cgn2\_6/ptodata/1/ina/6B\_COMB.seq:\*  
5: /cgn2\_6/ptodata/1/ina/PTUS\_COMB.seq:\*  
6: /cgn2\_6/ptodata/1/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	353	38.5	240	US-09-016-434-1028	Sequence 1028, Ap
2	184	17.9	1800	US-09-771-357-106	Sequence 106, App
3	154	16.8	1275	US-08-552-142A-3	Sequence 3, Appli
4	154	16.8	1275	US-08-910-973-3	Sequence 3, Appli
5	154	16.8	1275	US-09-499-227-3	Sequence 3, Appli
6	154	16.8	1275	PCT-US95-05741-3	Sequence 3, Appli
7	146.5	16.0	1560	US-08-552-142A-14	Sequence 14, Appl
8	146.5	16.0	1560	US-08-910-973-14	Sequence 14, Appl
9	146.5	16.0	1560	US-09-499-227-14	Sequence 14, Appl
10	146.5	16.0	2502	US-09-234-332-1	Sequence 1, Appli
11	146	15.9	2089	US-08-552-142A-1	Sequence 1, Appli
12	146	15.9	2089	US-08-910-973-1	Sequence 1, Appli

13 146 15.9 2089 4 US-09-499-227-1 Sequence 1, Appli  
14 146 15.9 2089 5 PCT-US95-05741-1 Sequence 1, Appli  
15 145.5 15.9 1462 1 US-08-552-142A-16 Sequence 16, Appli  
16 145 15.8 1951 1 US-08-910-973-16 Sequence 16, Appli  
17 145 15.8 1951 4 US-09-499-227-16 Sequence 16, Appli  
18 143.5 15.6 1527 4 US-08-722-570-12 Sequence 12, Appli  
19 143.5 15.6 1527 4 US-08-932-411A-12 Sequence 12, Appli  
20 143 15.6 1352 1 US-08-552-142A-10 Sequence 10, Appli  
21 143 15.6 1385 4 US-08-932-411A-17 Sequence 17, Appli  
22 143 15.6 1535 1 US-08-910-973-10 Sequence 10, Appli  
23 143 15.6 1535 4 US-09-499-227-10 Sequence 10, Appli  
24 143 15.6 1550 3 US-09-234-332-3 Sequence 3, Appli  
25 140.5 15.3 738 4 US-08-722-570-13 Sequence 13, Appli  
26 140.5 15.3 738 4 US-08-932-411A-13 Sequence 13, Appli  
27 140.5 15.3 1333 1 US-08-910-973-21 Sequence 21, Appli  
28 140.5 15.3 1333 4 US-09-499-227-21 Sequence 21, Appli  
29 140.5 15.3 1676 3 US-09-234-332-2 Sequence 2, Appli  
30 140 15.3 5340 4 US-09-535-145-1 Sequence 1, Appli  
31 138.5 15.1 524 1 US-08-552-142A-8 Sequence 8, Appli  
32 138.5 15.1 524 1 US-08-910-973-8 Sequence 8, Appli  
33 138.5 15.1 524 4 US-09-499-227-8 Sequence 8, Appli  
34 138.5 15.1 524 5 PCT-US95-05741-8 Sequence 8, Appli  
35 137 14.9 804 4 US-08-932-411A-19 Sequence 19, Appli  
36 137 14.9 1861 4 US-09-535-145-3 Sequence 3, Appli  
37 131 14.3 310 1 US-08-552-142A-12 Sequence 12, Appli  
38 129.5 14.1 485 5 PCT-US95-05741-10 Sequence 10, Appli  
39 126 13.7 1268 1 US-08-910-973-12 Sequence 12, Appli  
40 126 13.7 1268 4 US-09-499-227-12 Sequence 12, Appli  
41 125 13.6 1312 4 US-08-722-570-14 Sequence 14, Appli  
42 125 13.6 1312 4 US-08-932-411A-15 Sequence 15, Appli  
43 120 13.1 1238 4 US-09-774-528-350 Sequence 350, App  
44 119 13.0 1277 4 US-08-722-570-15 Sequence 15, Appli  
45 119 13.0 1277 4 US-08-932-411A-16 Sequence 16, Appli

#### ALIGNMENTS

##### RESULT 1

US-09-016-434-1028  
; Sequence 1028, Application US/09016434  
; Patent No. 6500938  
; GENERAL INFORMATION:  
; APPLICANT: Janice Au-Young  
; APPLICANT: Jeffrey J. Seilhamer  
; TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF SIGNALING  
; TITLE OF INVENTION: PATHWAY GENE EXPRESSION  
; NUMBER OF SEQUENCES: 1490  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.  
; STREET: 3174 PORTER DRIVE  
; CITY: PALO ALTO  
; STATE: CALIFORNIA  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/016,434  
; FILING DATE: HEREWITH  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Zeller, Karen J.  
; REGISTRATION NUMBER: 37,071  
; REFERENCE/DOCKET NUMBER: PA-0002 US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (650) 855-0555



```

; ORIGINAL SOURCE:
; ORGANISM: Xenopus laevis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 25..1083
US-08-552-142A-3
Alignment Scores:
Pred. No.: 3.64e-09 Length: 1275
Score: 154.00 Matches: 67
Percent Similarity: 46.46% Conservative: 25
Best Local Similarity: 33.84% Mismatches: 64
Query Match: 16.79% Indels: 43
DB: 1 Gaps: 8
US-09-701-674A-23 (1-179) x US-08-552-142A-3 (1-1275)
QY 5 SerLeuSerAspValGluAspLeuGlnGluValGluMetLeuGluCysAspGlyLeuLys 24
Db 103 AGTTCTCAGGATGAAACGATCTGGAGAA-----132
QY 25 MetAspSerAsnLysGluPheValThrSerAsnGluSerThrGluLysSerAsnCys 44
Db 133 -----AAGGAGGAGAGTGTGATGAAAGAGACGATGAAGACTCACTGAATCAT 180
QY 45 GluAsnGlySerProGlnLysGlyArgGly-----GlyLeuGlyLysArg---59
Db 181 CACATGGAGAGGA-GAACGAGGAGAGGATGAAGGGGATGAGGAGGAGGAGCATGA 239
QY 60 -----AAGGAGGAGAGTGTGATGAAAGAGACGATGAAGACTCACTGAATCAT 180
Db 240 AGATGATGATGAGGATGACGACCAACCCAGAGGAGGAGGAGGAGGAGGAGGAGGAG 239
QY 69 userGlyValSerGlnGluGlyLysGlnValGlnArgAsnAlaAlaAsnAlaArgGluAr 89
Db 300 GACGAAAGCCCGGTGGAGGATTTAAAGTGAAGCAGCATGAAGCAACCCAGGAGAG 359
QY 89 gAlaArgMetArgValLeuSerLysAlaPheSerArgLeuLysThrLeuProTirpVa 109
Db 360 GAATCGCATGACGGACTCAACGATCCCTGACAGCTGCGCAAGTGTGCTGCTGCTGCTA 419
QY 109 lProProAspThrLysLeuSerLysLeuAspThrLeuArgLeuAlaSerSerTyrIleAl 129
Db 420 CTCGAAACACAAAGTTGCTTAAGATTGAAACTCTGCGCTGCTGCTGCTGCTGCTGCTG 479
QY 129 ahLysLeuArgGlnIleLeuAlaAsnAspLysTyrGluAsn-----GlyTyrIleHisPr 147
Db 480 GGCTCTTCTGAGATTTAAGTTCGCGCAAGAACCCGACACCTGCTGCTGCTGCTGCTGCTG 539
QY 147 oVal-----AsnLeuThrTrpProPhe-----MetValAlaGly-----Ly 159
Db 540 TCTCTGCAAAAGTTTCTGCGAGCCACCAACCAATCTAGTAGCGGGGTGTCTGACGTGAA 599
QY 159 s-ProGlu---SerAspLeuLysGluValValThrAlaSerArgLeuCys 174
Db 600 CCCAGAACTTCTCTCTGAGAGTCCCTGAGCAGAGTCCAGACATCCAGTCCGACATGC 649
RESULT 4
US-08-910-973-3
; Sequence 3 Application US/06910973
; Patent No. 5795723
; GENERAL INFORMATION:
; APPLICANT: Tapscott, Stephen J.
; APPLICANT: Olson, James M.
; TITLE OF INVENTION: Expression of Neurogenic bHLH Genes in Primitive Neuroectoder
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Christensen O'Connor Johnson Kindness PLLC
; STREET: 1420 Fifth Avenue, Suite 2800
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98101-2347

```

```

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA: US/08/910,973
; APPLICATION NUMBER: US/08/910,973
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/239,238
; FILING DATE: 06-MAY-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/US95/05741
; FILING DATE: 08-MAY-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/17532
; FILING DATE: 30-October-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Sheiness, Diana K.
; REGISTRATION NUMBER: 35,356
; REFERENCE/DOCKET NUMBER: FPCR-1-10958
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-682-8100; 206-224-0735 (direct)
; TELEFAX: 206-225-0779
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1275 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; ORIGINAL SOURCE:
; ORGANISM: Xenopus laevis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 25..1083
US-08-910-973-3
Alignment Scores:
Pred. No.: 3.64e-09 Length: 1275
Score: 154.00 Matches: 67
Percent Similarity: 46.46% Conservative: 25
Best Local Similarity: 33.84% Mismatches: 64
Query Match: 16.79% Indels: 43
DB: 1 Gaps: 8
US-09-701-674A-23 (1-179) x US-08-910-973-3 (1-1275)
QY 5 SerLeuSerAspValGluAspLeuGlnGluValGluMetLeuGluCysAspGlyLeuLys 24
Db 103 AGTTCTCAGGATGAAACGATCTGGAGAA-----132
QY 25 MetAspSerAsnLysGluPheValThrSerAsnGluSerThrGluLysSerAsnCys 44
Db 133 -----AAGGAGGAGAGTGTGATGAAAGAGACGATGAAGACTCACTGAATCAT 180
QY 45 GluAsnGlySerProGlnLysGlyArgGly-----GlyLeuGlyLysArg---59
Db 181 CACATGGAGAGGA-GAACGAGGAGAGGATGAAGGGGATGAGGAGGAGGAGGAGGAGGAG 239
QY 60 -----AAGGAGGAGAGTGTGATGAAAGAGACGATGAAGACTCACTGAATCAT 180
Db 240 AGATGATGATGAGGATGACGACCAACCCAGAGGAGGAGGAGGAGGAGGAGGAGGAG 239
QY 69 userGlyValSerGlnGluGlyLysGlnValGlnArgAsnAlaAlaAsnAlaArgGluAr 89
Db 300 GACGAAAGCCCGGTGGAGGATTTAAAGTGAAGCAGCATGAAGCAACCCAGGAGAG 359
QY 89 gAlaArgMetArgValLeuSerLysAlaPheSerArgLeuLysThrLeuProTirpVa 109
Db 360 GAATCGCATGACGGACTCAACGATCCCTGACAGCTGCGCAAGTGTGCTGCTGCTGCTA 419

```

QY 109 lProProAspThrLysLeuSerLysLeuAspThrLeuArgLeuAlaSerSerTyrIleAl 129  
 Db 420 CTCACAAACACAAAGTTGCTAAGATTGAACCTTCGCCCTGGCTAAGAACTACATCG 479  
 QY 129 aHisLeuArgGlnIleLeuAlaSerLysTyrGluAsn-----GlyTyrIleHisPr 147  
 Db 480 GGCTCTTTCTGAGATTAAAGTTCGGCAAAAGCCAGACCTGGTGTCTCTTTGTACAAAC 539  
 QY 147 oVal-----AsnLeuThrTrpProPhe-----MetValAlaGly-----Ly 159  
 Db 540 TCTCTGCAAAAGTTTTCGACGCCACCCACCAATCTAGTAGCGGGGTGCTCGACCTGAA 599  
 QY 159 s-ProGlu---SerAspLeuLysGluValValThrAlaSerArgLeuCys 174  
 Db 600 CCCAGAACTTTCCTCTGAGCAGAGTCAGGACATCCAGTCGCACATGC 649

## RESULT 5

US-09-499-227-3  
 ; Sequence 3, Application US/09499227  
 ; Patent No. 6444463  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Tapscott, Stephen J.  
 ; APPLICANT: Olson, James M.  
 ; TITLE OF INVENTION: Expression of Neurogenic bHLH Genes in Primitive Neuroectoderm  
 ; NUMBER OF SEQUENCES: 24  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Christensen O'Connor Johnson Kindness PLLC  
 ; STREET: 1420 Fifth Avenue, Suite 2800  
 ; CITY: Seattle  
 ; STATE: WA  
 ; COUNTRY: USA  
 ; ZIP: 98101-2347  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patent in Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/09/499,227  
 ; FILING DATE: 05-August-1998  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 08/239,238  
 ; FILING DATE: 06-May-1994  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: WO PCT/US95/05741  
 ; FILING DATE: 08-May-1995  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: PCT/US96/17532  
 ; FILING DATE: 30-October-1996  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 08/910,973  
 ; FILING DATE: 07-August-1997  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Sheiness, Diana K.  
 ; REGISTRATION NUMBER: 35,356  
 ; REFERENCE/DOCKET NUMBER: FHC8-1-12742  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 206-682-8100; 206-224-0735 (direct)  
 ; TELEFAX: 206-225-0779  
 ; INFORMATION FOR SEQ ID NO: 3:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 1275 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: double  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: cDNA  
 ; ORIGINAL SOURCE:  
 ; ORGANISM: Xenopus laevis  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: 25..1083  
 ; US-09-499-227-3

Alignment Scores:  
 Pred. No.: 3,64e-09 Length: 1275  
 Score: 154.00 Matches: 67  
 Percent Similarity: 46.46% Conservative: 25  
 Best Local Similarity: 33.84% Mismatches: 64  
 Query Match: 16.79% Indels: 43  
 DB: 8  
 Gaps: 8

US-09-701-674A-23 (1-179) x US-09-499-227-3 (1-1275)

QY 5 SerLeuSerAspValGluAspLeuGlnGluValGluMetLeuGluCysAspGlyLeuLys 24  
 Db 103 AGTTCTCAGGATGAAAACGATCTCGAGAA----- 132  
 QY 25 MetAspSerAsnLysGluPheValThrSerAsnGluSerThrGluGluSerSerAsnCys 44  
 Db 133 -----AAGGAGGAGAGATTGATGAAGAAGACGATGAAGACTCAGTGAATCAT 180  
 QY 45 GluAsnGlySerProGlnLysGlyArgGly-----GlyLeuGlyLysArg--- 59  
 Db 181 CACAATCGCAGAGGA-GAACGAGGAGAGGATGAAGGGGATGAGGAGGAGGAGGAGCATGA 239  
 QY 60 -----ArgLysAlaProThr-LysLysSerProLe 69  
 Db 240 AGATGATGATGAGATGACACGACAGAAACCCAAAGCGGAGGACCGAAAGAAAAAAT 299  
 QY 69 userGlyValSerGlnGluGlyLysGlnValGlnArgAsnAlaAlaAsnAlaArgGluAr 89  
 Db 300 GACGAAAGCCCGGTGGAGCGATTAAAGTGAGACGCATGAAGGCCAAACGCCCAGGAGAG 359  
 QY 89 gAlaArgMetArgValLeuSerLysAlaPheSerArgLeuLysThrLeuProTrpVa 109  
 Db 360 GAATCGCATGCGGACTCAACGATCCCTGGACACTCTCGCAAAAGTTGTGCCCTGCTA 419  
 QY 109 lProProAspThrLysLeuSerLysLeuAspThrLeuArgLeuAlaSerSerTyrIleAl 129  
 Db 420 CTCACAAACACAAAGTTGCTAAGATTGAACCTTCGCCCTGGCTAAGAACTACATCTG 479  
 QY 129 aHisLeuArgGlnIleLeuAlaSerLysTyrGluAsn-----GlyTyrIleHisPr 147  
 Db 480 GGCTCTTTCTGAGATTAAAGTTCGGCAAAAGCCAGACCTGGTGTCTCTTTGTACAAAC 539  
 QY 147 oVal-----AsnLeuThrTrpProPhe-----MetValAlaGly-----Ly 159  
 Db 540 TCTCTGCAAAAGTTTTCGACGCCACCCACCAATCTAGTAGCGGGGTGCTCGACCTGAA 599  
 QY 159 s-ProGlu---SerAspLeuLysGluValValThrAlaSerArgLeuCys 174  
 Db 600 CCCAGAACTTTCCTCTGAGCAGAGTCAGGACATCCAGTCGCACATGC 649

## RESULT 6

PCT-US95-05741-3  
 ; Sequence 3, Application PC/TUS9505741  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Weintraub, Harold  
 ; APPLICANT: Lee, Jacqueline B.  
 ; APPLICANT: Tapscott, Stephen J.  
 ; APPLICANT: Hollenberg, Stanley M.  
 ; TITLE OF INVENTION: Neurogenic Differentiation (Neurod) Gene  
 ; NUMBER OF SEQUENCES: 11  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Christensen O'Connor Johnson Kindness  
 ; STREET: 1420 Fifth Avenue, Suite 2800  
 ; CITY: Seattle  
 ; STATE: WA  
 ; COUNTRY: USA  
 ; ZIP: 98101-2347  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patent in Release #1.0, Version #1.25

/ CURRENT APPLICATION DATA:  
 / APPLICATION NUMBER: PCT/US95/05741  
 / FILING DATE:  
 / CLASSIFICATION:  
 / ATTORNEY/AGENT INFORMATION:  
 / NAME: Broderick, Thomas F.  
 / REGISTRATION NUMBER: 31,332  
 / REFERENCE/DOCKET NUMBER: FPCR-1-8504  
 / TELECOMMUNICATION INFORMATION:  
 / TELEPHONE: 206-682-8100  
 / TELEFAX: 206-225-0709  
 / INFORMATION FOR SEQ ID NO: 3:  
 / SEQUENCE CHARACTERISTICS:  
 / LENGTH: 1275 base pairs  
 / TYPE: nucleic acid  
 / STRANDEDNESS: double  
 / TOPOLOGY: linear  
 / MOLECULE TYPE: cDNA  
 / ORIGINAL SOURCE:  
 / ORGANISM: Xenopus laevis  
 / FEATURE:  
 / NAME/KEY: CDS  
 / LOCATION: 25..1083  
 / PCT-US95-05741-3

Alignment Scores:  
 Pred. No.: 3,64e-09 Length: 1275  
 Score: 154.00 Matches: 67  
 Percent Similarity: 46.46% Conservative: 25  
 Best Local Similarity: 33.84% Mismatches: 64  
 Query Match: 16.79% Indels: 43  
 DB: 5 Gaps: 8

US-09-701-674A-23 (1-179) x PCT-US95-05741-3 (1-1275)

QY	5	SeuSerAspValGluAspLeuGluValGluMetLeuGluCysAspGlyLeuLys	24
Db	103	AGTTCTCAGGAGAAACGATCTGGAGAA	132
QY	25	MetAspSerAsnLysGluPheValThrSerAsnGluSerThrGluGluSerSerAsnCys	44
Db	133	-----AAGGAGGAGAGTGTGATGAAGAAGACGATGAAGACTCATCTGAATCAT	180
QY	45	GluAsnGlySerProGlnLysGlyValGly-----GlyLeuGlyLysArg	59
Db	181	CACATGAGAGGA-GAAGGAGAGGATGAAGGGGATGAGGAGGAGGAGCATGA	239
QY	60	-----ArgLysAlaProThr-LysLysSerProLe	69
Db	240	AGATGATGATGAGGATGACGACCAAGAAACCCAAAGGCGAGGACCGAAAGAAAAAAT	299
QY	69	uSerGlyValSerGluGlyLysGlnValGlnArgAsnAlaAlaAsnAlaArgGluAr	89
Db	300	GACGAAGCCGGTGGAGGATTTAAGTGAGACCATGAAGGCCAAACCCAGGAGAG	359
QY	89	GAlaArgMetArgValLeuSerLysAlaPheSerArgLeuLysThrThrLeuProTrpVa	109
Db	360	GAATCGCATGACGCGACTCAACGATCCCTGGACAGTCTGCGCAAAAGTTGTGCTCGCTA	419
QY	109	lProProAspThrLysLeuSerLysLeuAspThrLeuArgLeuAlaSerSerTyxIleAl	129
Db	420	CTCAAAACACAAAAGTTGTCTAAGATTGAAATCTGCGCTGGCTGAAAGATCATCTG	479
QY	129	aHisLeuArgGlnIleLeuAlaAsnAspLysTyxGluAsn-----GlyTyxIleHisPr	147
Db	480	GGCTCTTTCTGAGATTTTAAGTTCGCGCAAAAGCCAGACCTGGTCTCTTTGTACAAAC	539
QY	147	oVal-----AsnLeuThrTrpProPhe-----MetValAlaGly-----Ly	159
Db	540	TCCTCTCAAAAGTTGTCTGACGACCCACCAATCTAGTAGCGGGTGTCTCAGCTGAA	599
QY	159	s-ProGlu---SerAspLeuLysGluValValThrAlaSerArgLeuLys	174

Db 600 CCCAGAACTTCTCTCTGAGCAGAGTCAGGACATCCAGTCGACATGC 649  
 RESULT 7  
 US-08-552-142A-14  
 / Sequence 14: Application US/08552142A  
 / Patent No.: 569395  
 / GENERAL INFORMATION:  
 / APPLICANT: Weintraub, Harold M.  
 / APPLICANT: Lee, Jacqueline E.  
 / APPLICANT: Tapscott, Stephen J.  
 / APPLICANT: Hollenberg, Stanley M.  
 / TITLE OF INVENTION: Neurogenic Differentiation (NeuroD) Genes  
 / TITLE OF INVENTION: and Proteins  
 / NUMBER OF SEQUENCES: 20  
 / CORRESPONDENCE ADDRESS:  
 / ADDRESSEE: Christensen O'Connor Johnson Kindness PLLC  
 / STREET: 1420 Fifth Avenue, Suite 2800  
 / CITY: Seattle  
 / STATE: WA  
 / COUNTRY: USA  
 / ZIP: 98101-2347  
 / COMPUTER READABLE FORM:  
 / MEDIUM TYPE: Floppy disk  
 / COMPUTER: IBM PC compatible  
 / OPERATING SYSTEM: PC-DOS/MS-DOS  
 / SOFTWARE: PatentIn Release #1.0, Version #1.25  
 / CURRENT APPLICATION DATA:  
 / APPLICATION NUMBER: US/08/552,142A  
 / FILING DATE: 02-NOV-1995  
 / CLASSIFICATION: 514  
 / PRIOR APPLICATION DATA:  
 / APPLICATION NUMBER: US 08/239,238  
 / FILING DATE: 06-MAY-1994  
 / PRIOR APPLICATION DATA:  
 / APPLICATION NUMBER: WO PCT/US95/05741  
 / FILING DATE: 08-MAY-1995  
 / ATTORNEY/AGENT INFORMATION:  
 / NAME: Broderick, Thomas F.  
 / REGISTRATION NUMBER: 31,332  
 / REFERENCE/DOCKET NUMBER: FPCR-1-8933  
 / TELECOMMUNICATION INFORMATION:  
 / TELEPHONE: 206-682-8100  
 / TELEFAX: 206-225-0709  
 / INFORMATION FOR SEQ ID NO: 14:  
 / SEQUENCE CHARACTERISTICS:  
 / LENGTH: 1560 base pairs  
 / TYPE: nucleic acid  
 / STRANDEDNESS: double  
 / TOPOLOGY: linear  
 / MOLECULE TYPE: cDNA  
 / ORIGINAL SOURCE:  
 / ORGANISM: Homo sapiens  
 / IMMEDIATE SOURCE:  
 / CLONE: HC2A  
 / FEATURE:  
 / NAME/KEY: CDS  
 / LOCATION: 57..1126  
 / US-08-552-142A-14  
 Alignment Scores:  
 Pred. No.: 4,16e-08 Length: 1560  
 Score: 146.50 Matches: 41  
 Percent Similarity: 50.37% Conservative: 27  
 Best Local Similarity: 30.37% Mismatches: 66  
 Query Match: 15.98% Indels: 1  
 DB: 1 Gaps: 1

US-09-701-674A-23 (1-179) x US-08-552-142A-14 (1-1560)

QY	6	LeuSerAspValGluAspLeuGluValGluMetLeuGluCysAspGlyLeuLysMet	25
Db	135	CTCAGTTCTCAGGAGGAGGACGACGACACCAAGAGGAGGACGACCTCGAAGCCATG	194







ADDRESSEE: Christensen O'Connor Johnson Kindness PLLC  
STREET: 1420 Fifth Avenue, Suite 2800  
CITY: Seattle  
STATE: WA

COUNTRY: USA

ZIP: 98101-2347

COMPUTER TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/552.142A

FILING DATE: 02-NOV-1995

CLASSIFICATION: 514

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/239,238

FILING DATE: 06-MAY-1994

PRIOR APPLICATION DATA:

APPLICATION NUMBER: WO PCT/US95/05741

FILING DATE: 08-MAY-1995

ATTORNEY/AGENT INFORMATION:

NAME: Broderick, Thomas F.

REGISTRATION NUMBER: 31,332

REFERENCE/DOCKET NUMBER: PHCR-1-8933

TELECOMMUNICATION INFORMATION:

TELEPHONE: 206-682-8100

TELEFAX: 206-225-0709

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 2089 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: cdna

ORGANISM: Mus musculus

FEATURE:

NAME/KEY: CDS

LOCATION: 229..1302

US-08-552-142A-1

Alignment Scores:  
Pred. No.: 7,49e-08 Length: 2089  
Score: 146.00 Matches: 45  
Percent Similarity: 49.23% Conservative: 19  
Best Local Similarity: 34.62% Mismatches: 58  
Query Match: 15.92% Indels: 8  
DB: 1 Gaps: 3

US-09-701-674a-23 (1-179) x US-08-552-142A-1 (1-2089)

Qy 10 GluAspLeuGlnGluValGluMetLeuGluCysAspGlyLeuLysMetAspSerAsnLys 29  
Db 346 GAGGACGAGTGAAGCCATGAATGACGAGGAGGAGTCTCTGAGA---AACGGGGGAGAG 402  
Qy 30 GluPheValThrSerAsnGluSerThrGluGluSerSerAsnCysGluAsnGlySerPro 49  
Db 403 GAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 462  
Qy 50 GlnLysGlyArgGlyGlyLeuGlyLysArgArgLysAlaProThrLysLysSerProLeu 69  
Db 463 CAAAAG-----CCCAAGAGACGGGGTCCC---AAAAAGAAAAGATG 501  
Qy 70 SerGlyValSerGlnGluGlyLysGlnValGlnArgAsnAlaAlaAsnAlaArgGluArg 89  
Db 502 ACCAAGCGCGCTAGACGCTTTTAATTAAGGCGCATGAAGCCCAACCCCGGAGCGG 561  
Qy 90 AlaArgMetArgValLeuSerLysAlaPheSerArgLeuLysThrThrLeuProTrpVal 109  
Db 562 AACCGCATGCACGGCTGACGGCGCTGGACAACTGCGCAAGTGGTGTACCTTGTCTAC 621  
Qy 110 ProProAspThrLysLeuSerLysLeuAspThrLeuArgLeuAlaSerSerTyrlleAla 129

Db 622 TCCAAGACCCAGAAACTGTCTAAATAGACACTGCGCTTGCCCAAGAACTACATCTGG 681

Qy 130 HisLeuArgGlnIleLeuAlaAsnAspLys 139

Db 682 GCTCTGTGACAGATCTCTGCGCTCAGGCAAA 711

RESULT 12

US-08-910-973-1

Sequence 1, Application US/08910973

Patent No. 5795723

GENERAL INFORMATION:

APPLICANT: Tapscott, Stephen J.

APPLICANT: Olson, James M.

TITLE OF INVENTION: Expression of Neurogenic bHLH Genes in Primitive Neuroectoder

NUMBER OF SEQUENCES: 24

CORRESPONDENCE ADDRESS:

ADDRESSEE: Christensen O'Connor Johnson Kindness PLLC

STREET: 1420 Fifth Avenue, Suite 2800

CITY: Seattle

STATE: WA

COUNTRY: USA

ZIP: 98101-2347

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/910,973

FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/239,238

FILING DATE: 06-MAY-1994

PRIOR APPLICATION DATA:

APPLICATION NUMBER: WO PCT/US95/05741

FILING DATE: 08-MAY-1995

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/US96/17532

FILING DATE: 30-October-1996

ATTORNEY/AGENT INFORMATION:

NAME: Sheiness, Diana K.

REGISTRATION NUMBER: 35,356

REFERENCE/DOCKET NUMBER: PHCR-1-10958

TELECOMMUNICATION INFORMATION:

TELEPHONE: 206-682-8100; 206-224-0735 (direct)

TELEFAX: 206-225-0779

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 2089 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: cdna

ORGANISM: Mus musculus

FEATURE:

NAME/KEY: CDS

LOCATION: 229..1302

US-08-910-973-1

Alignment Scores:

Pred. No.: 7,49e-08 Length: 2089  
Score: 146.00 Matches: 45  
Percent Similarity: 49.23% Conservative: 19  
Best Local Similarity: 34.62% Mismatches: 58  
Query Match: 15.92% Indels: 8  
DB: 1 Gaps: 3

US-09-701-674A-23 (1-179) x US-08-910-973-1 (1-2089)

Qy 10 GluAspLeuGlnGluValGluMetLeuGluCysAspGlyLeuLysMetAspSerAsnLys 29





```

Qy 143 -----GlyTyrIleHisProVal-AsnLeuThrTrpProPheMetV 156
Db 797 CGTGCAGACTCTGTGCAAGGGGCTGTACAGAGCCACCGAATCTGGTGGCGGCTGCCT 856
Qy 156 alAlaGlyLysProGluSerAspLeuLysGluValValThrAlaSerArgLeuCysGlyT 176
Db 857 GCAGT-----TAACTCTCGTAACTTCTCTCAGGAGCGGGCC 895
Qy 176 hrThrAlaSer 179
Db 896 GGACGGCGGCC 906

```

Search completed: December 17, 2004, 00:05:29  
 Job time : 94 secs

This Page Blank (uspto)

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: December 16, 2004, 20:16:52 ; Search time 455 seconds

(without alignments)

2065.157 Million cell updates/sec

Title: US-09-701-674a-23

Perfect score: 917

Sequence: 1 MFTGSLSDVEDLQEVMELEC.....PESDLKVVVTASRLCGTTAS 179

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 4134886 segs, 2624710521 residues

Total number of hits satisfying chosen parameters: 8269772

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh  
-Q=/cgn2\_1/USPTO.spool\_h/US09701674/runat\_15122004\_100440\_27494/app\_query.fasta\_1.327  
-DB=N Geneseq 23Sep04 -QFMT=fastap -SUFFIX=ring -MINMATCH=0.1 -LOOPEL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR SCORE=prt -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEA=SIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09701674 -CGN\_1\_1\_470 -runat\_15122004\_100440\_27494 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N Geneseq 23Sep04: \*  
1: Geneseqn1980s: \*  
2: Geneseqn1990s: \*  
3: Geneseqn2000s: \*  
4: Geneseqn2001as: \*  
5: Geneseqn2001bs: \*  
6: Geneseqn2002as: \*  
7: Geneseqn2002bs: \*  
8: Geneseqn2003as: \*  
9: Geneseqn2003bs: \*  
10: Geneseqn2003cs: \*  
11: Geneseqn2003ds: \*  
12: Geneseqn2004s: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	917	100.0	1272	3	Aaz57861
2	764.5	83.4	632	6	ABQ55427 Human ova
3	514	56.1	2177	10	ABQ25693 Human cdn
4	514	56.1	2196	3	AAJ72428 Human nuc
5	514	56.1	2196	10	ADJ56314 Human cdn
6	507	55.3	2382	6	AAS94828 Human DNA

7	495.5	54.0	1718	3	AAZ50464
8	493.5	53.8	1716	3	AAZ50465
9	487	53.1	482	9	ACH21857
10	353	38.5	240	10	ACA56430
11	353	38.5	240	12	ADJ56226
12	318	34.7	697	6	ABL65519
13	318	34.7	697	6	ABL65935
14	318	34.7	697	6	ABK64784
15	228	24.9	493	9	ACH23969
16	211.5	23.1	1065	4	ABL08723
17	208	22.7	5101	4	ABL08722
18	185	20.2	963	8	ABZ18589
19	183.5	20.0	1276	10	ADC30883
20	183.5	20.0	1395	12	ADQ25304
21	183.5	20.0	1406	8	ACC62520
22	183.5	20.0	1406	8	ACC62491
23	180.5	19.7	1140	6	ABQ88217
24	180.5	19.7	1160	10	ADC32623
25	174.5	19.0	981	10	ADD51583
26	169	18.4	474	5	AAS76438
27	168.5	18.4	1004	6	ABK71581
28	165.5	18.0	651	4	AAH48230
29	165.5	18.0	651	4	AAH49596
30	165.5	18.0	651	4	AAH44361
31	165.5	18.0	2466	8	ABX63251
32	164	17.9	609	12	ACH89545
33	164	17.9	797	11	ADP65591
34	164	17.9	1396	11	ADP65052
35	164	17.9	1664	12	ADQ22344
36	164	17.9	1800	6	ABT06509
37	164	17.9	1800	11	ADM83716
38	164	17.9	2870	12	ADQ17684
39	162	17.7	531	4	ABL16071
40	162	17.7	1457	6	ABL63712
41	162	17.7	1457	11	ADP65592
42	162	17.7	1678	6	AAS94864
43	157.5	17.2	907	9	AAI62234
44	157.5	17.2	2537	6	ABSS1803
45	157.5	17.2	2550	10	ADD29657

## ALIGNMENTS

RESULT 1

AAZ57861

ID AAZ57861 standard; cdna; 1272 BP.

XX AC AAZ57861;

XX AC AAZ57861;

DT 11-APR-2000 (first entry)

XX DE Protein regulating gene expression PRGE-23 cDNA clone 3340296.

XX DE Protein regulating gene expression; PRGE-23; human; cancer; inflammation;

XX KW anticancer; antitumor; antiinflammatory; Myc; HLH protein; gene therapy;

XX KW diagnosis; ss.

XX OS Homo sapiens.

XX FH Key

FT CDS Location/Qualifiers

FT CDS 268..807

FT CDS /\*tag= a

XX PN WO9964596-A2.

XX PD 16-DEC-1999.

XX PF 11-JUN-1999;

XX PR 12-JUN-1998;

XX PR 29-JUL-1998;

XX PR 14-OCT-1998;

XX PR 98US-0089029P.

XX PR 98US-0094575P.

XX PR 98US-0104624P.

XX PR 99WO-US013281.



PA (INCY-) INCYTE PHARM INC.  
XX Lall P, Yue H, Tang YT, Hillman JL, Bandman O, Corley NC;  
PI Guegler KJ, Gorgone GA, Baughn WR, Patterson C, Lu DM;  
XX WPI; 2000-116543/10.  
DR P-PSDB; AAY58630.  
XX New human polypeptides that regulate gene expression, for treatment,  
PT prevention and diagnosis of, e.g. cancer.  
XX Claim 9; Page 139; 150pp; English.  
XX The present sequence is that of Incyte clone 3340296 encoding new human  
CC protein regulating gene expression PRGE-23 (see AAY58630). The cDNA was  
CC initially isolated from spleen tissue cDNA library SPLNNOT10, and the  
CC full-length sequence assembled from overlapping clones from a number of  
CC libraries. PRGE-23 is expressed in reproductive, developmental and  
CC urologic tissues associated with cancer, inflammation and foetal  
CC diseases, disorders or conditions. It is characterised as an Myc-type HLH  
CC protein. The invention provides PRGE polypeptides (see AAY58608-38) and  
CC polynucleotides (see AAZ57839-69), expression vectors, host cells,  
CC antibodies, agonists and antagonists. It also provides methods for  
CC diagnosing, treating or preventing disorders associated with expression  
CC of PRGE. Polynucleotides are also used as sources of probes and primers  
CC for diagnosis and monitoring of disease, also for detecting related  
XX sequences and in gene mapping  
XX SQ Sequence 1272 BP; 311 A; 358 C; 318 G; 285 T; 0 U; 0 Other;  
Alignment Scores:  
Pred. No.: 4,19e-92 Length: 1272  
Score: 917.00 Matches: 179  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 3 Gaps: 0  
US-09-701-674A-23 (1-179) x AAZ57861 (1-1272)  
QY 1 MetSerThrGlySerLeuSerAspValGluAppLeuGlnGluValMetLeuGluCys 20  
DB 268 ATGTCACCGGCTCCCTCAGCGATGTTGGAGGACCTTCAAGAGGTGGAGATGTGGAATGT 327  
QY 21 AspGlyLeuMethSerAspSerLeuGluPheValThrSerAsnGluSerThrGluGlu 40  
DB 328 GACGGGTTGAAATGATTCGAACACAGGAATTTGTGACTTCCACGAGACGCCGAGGAG 387  
QY 41 SerSerAsnCysGluAsnGlySerProGlnLysGlyArgGlyLeuGlyLysArgArg 60  
DB 388 AGCTCCAACTGCGAGAAATGGTCTCCAGAGAGGGCGCGCGCTGGCGCAAGAGGAGG 447  
QY 61 LysAlaProThrLysLysSerProLeuSerGlyValSerGlnGluGlyGlnValGln 80  
DB 448 AAGGCGCCCAACCAAGAGAGCCCTTGGCGGGGTTCAGCCAGGAGGAGGAGGAGTCCAG 507  
QY 81 ArgAsnAlaAlaAsnAlaArgGluArgAlaArgMetArgValLeuSerLysAlaPheSer 100  
DB 508 CGCAACGCGCCCAACGCGGAGAGCGGGCCCGCATGCGAGTCTGAGCAAGGCCCTTCTCC 567  
QY 101 ArgLeuLysThrThrLeuProThrValProProAspThrLysLeuSerLysLeuAspThr 120  
DB 568 AGACTCAAGACCAACCTTGGCTGGTGGCGCCCGCCGACCAACCAAGCTCTCCAAAGTGGACACG 627  
QY 121 LeuArgLeuAlaSerSerThrLeuAlaHisLeuArgGlnLeuAlaAsnAspLysTyr 140  
DB 628 CTGAGGCTGGCGTCCAGCTACATCGCCCACTTGAGGCGAGATCTCTGCTTACCAACCAATAC 687  
QY 141 GluAsnGlyTyrIleHisProValAsnLeuThrTrpProPheMetValAlaGlyLysPro 160  
DB 688 GAGACGGGGTACATTACCCGGTCAACCTGACGCTGGCCCTTTATGTTGGTCCCGGGAACCC 747  
QY 161 GluSerAspLeuLysGluValValThrAlaSerArgLeuCysGlyThrThrAlaSer 179  
Db 748 GAGAGTGACCTGAAGAAGTGTGACCGGAGCGCTTATGTGGACCAACCGCGTCC 804  
RESULT 2  
ABQ55427  
ID ABQ55427 standard; cDNA; 632 BP.  
XX ABQ55427;  
XX 22-AUG-2002 (first entry)  
XX Human ovarian antigen HNOJ185 cDNA, SEQ ID NO:1307.  
XX Human; ovarian antigen; ovary; ovarian; breast; cancer; tumour;  
KW ovarian cancer; breast cancer; tumour; reproductive system disorder;  
KW infertility; pregnancy disorder; anovulation; polycystic ovary syndrome;  
KW PCOS; ovarian cyst; dysmenorrhea; endocrine disorder; infection;  
KW inflammatory condition; immune disorder; blood disorder;  
KW cardiovascular disorder; respiratory disorder; neurological disorder;  
KW gastrointestinal disorder; urinary system disorder; drug screening;  
KW gene therapy; chromosome mapping; forensic analysis;  
KW antibody preparation; cytostatic; immunomodulatory; neuroprotective;  
KW antiinflammatory; gynaecological; reproductive; gene; ss.  
OS Homo sapiens.  
XX WO200200677-A1.  
XX 03-JAN-2002.  
XX 07-JUN-2001; 2001WO-US018569.  
XX 07-JUN-2000; 2000US-0209467P.  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX Birse CE, Rosen CA;  
XX WPI; 2002-147878/19.  
XX P-PSDB; ABP42350.  
XX Isolated nucleic acid molecules encoding novel ovarian polypeptides,  
PT useful in the prevention, treatment and diagnosis of cancer (e.g. ovarian  
PT cancer), immune disorders, cardiovascular disorders and neurological  
PT diseases.  
XX Claim 1; SEQ ID NO 1307; 2922pp; English.  
XX The invention relates to 2175 novel human ovarian antigens (ABP41054-  
CC ABP43228) and to cDNAs encoding them (ABQ54131-ABQ56305), and also  
CC encompasses polypeptides 90% identical and polynucleotides 95% identical  
CC to the sequences of the invention. The invention additionally relates to  
CC recombinant vectors and host cells comprising human ovarian antigen  
CC polynucleotides, antibodies against human ovarian antigens, and the use  
CC of ovarian antigen polynucleotides and polypeptides in diagnosing,  
CC treating, prognosing or preventing various ovary and/or breast-related  
CC disorders. Such conditions include ovarian cancer and breast cancer, and  
CC metastatic tumours of ovarian or breast origin, reproductive system  
CC disorders (e.g., infertility, disorders of pregnancy, anovulation,  
CC polycystic ovary syndrome, ovarian cysts, and dysmenorrhea), endocrine  
CC disorders, infections (e.g., chlamydia, HIV, toxoplasmosis, and toxic  
CC shock syndrome), inflammatory conditions (e.g., mastitis, oophoritis and  
CC vaginitis), immune disorders (e.g., congenital and acquired  
CC immunodeficiencies, autoimmune oophoritis, systemic lupus erythematosus),  
CC blood-related disorders (e.g., anaemia), cardiovascular disorders,  
CC respiratory disorders, neurological disorders, gastrointestinal disorders  
CC and urinary system disorders. Ovarian antigen polypeptides and  
CC polynucleotides may also be used in screening for compounds which  
CC modulate ovarian antigen expression or activity. The polynucleotides may  
CC further be used for gene therapy, chromosome mapping, in the  
CC identification of individuals and in forensic analysis; and the  
CC polypeptides may be used as food additives or to prepare antibodies  
CC useful in disease diagnosis, drug targeting and phenotyping. The present

CC sequence represents cDNA encoding a human ovarian antigen of the  
CC invention. Note: The sequence data for this patent did not form part of  
CC the printed specification, but was obtained in electronic format directly  
CC from WIPO at [ftp.wipo.int/pub/published\\_pct\\_sequences](http://ftp.wipo.int/pub/published_pct_sequences)  
XX  
SQ Sequence 632 BP; 145 A; 176 C; 201 G; 93 T; 0 U; 17 Other;

Alignment Scores:	1.56e-75	Length:	632
Pred. No.:	744.50	Matches:	163
Score:	91.62%	Conservative:	1
Percent Similarity:	91.06%	Mismatches:	14
Best Local Similarity:	83.37%	Indels:	3
Query Match:		Gaps:	0
DB:	6		

US-09-701-674A-23 (1-179) x ABQ55427 (1-632)

Qy	1	MetSerThrGlySerLeuSerAspValGluAspLeuGlnGluValGluMetLeuGluCys	20
Db	13	ATGTCACCGGCTCCCTCAGCGATGTGAGAGACCTTCAAGAGGTGGAGATGTTGGAAATGT	72
Qy	21	AspGlyLeuLysMetAspSerAsnLysGluPheValThrSerAsnGluSerThrGluGlu	40
Db	73	GACGGNTCAAAATGGATTGGAAACAGAAATTTGTGACTTCCACGAGAGCACCCGAGGAG	132
Qy	41	SerSerAsnCysGluAsnGlySerProGlnLysGlyArgGlyGlyLeuGlyLysArgArg	60
Db	133	AGCTCCAACTGGCGAGAAATGGGTCTCTCCACAGAGGGCGCGNGGCTTGGCGAAGAGGAG	192
Qy	61	LysAlaProThrLysLysSerProLeuSerGlyValSerGlnGluGlyLysGlnValGln	80
Db	193	ANGGCGCCCAACAGAAAGCCCCCTGAGCGGTGTCAAGCAGAGAGGGGAAGCAGGTCNAG	252
Qy	81	ArgAsnAlaAlaAsnAlaArgGluArgAlaArgMetArgValLeuSerLysAlaIpheSer	100
Db	253	CGCAACCGCGNACAGCGCGAGAGCGGNCCTGGCATGCGAGTGTGTAGCAAGGCGCTTCTCC	312
Qy	101	ArgLeuLysThrThrLeuProTirValProProAspThrLysLeuSerLysLeuAspThr	120
Db	313	AGACTCAAGACCAACCTGNCTGGGTGGCTCCCGCCCGACCAACAGCTCTCCAGCTGGACAG	372
Qy	121	LeuArgLeuAlaSerSerTyrIleAlaHisLeuArgGlnIleLeuAlaAsnAspLysTyr	140
Db	373	CTCAGGCTGGGGTCCAGCTACATCGCCACTTGAGGCAGATCTGTGCTAACGACCAATAC	432
Qy	141	GluAsnGlyTyrIleHisProValAsnLeuThrTirProPheMetValAlaGlyLysPro	160
Db	433	NAGAACGGGTACATTA--CCCGGGCACTGACGTGGNCTTTTATGTGGC-GGGAATCCC	489
Qy	161	GluSerAspLeuLysGluValVal-ThrAlaSerArgLeuCysGlyThrThrAla	178
Db	490	GAGAGTGANCTCAAGAAGTGGTTCACGGCAGCGCGTATTGTGGAAACACGCG	544

### RESULT 3

ADE25693  
 ID ADE25693 standard; cDNA; 2177 BP.

AC ADE25693:

XXXXXX

DT 29-JAN-2004 (first entry)

XX DE Human cDNA differentially expressed in foam cells #97.

Human; ss; differential expression; foam cell; LPS; lipopolysaccharide; cardiovascular disease; atherosclerosis.

Homo sapiens.

XX PN US2003194721-A1.

XX  
XX

PD 16-OCT-2003.

XX  
18-SEP-2002: 2002US-00247671.

XX PR XX PA XX PI XX DR DR XX PT PT PT PT XX PS XX CC

XX  
SQ Sequence 2177 BP; 423 A; 691 C; 650 G; 413 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	6.99e-47	Length:	2177
Score:	514.00	Matches:	113
Percent Similarity:	63.21%	Conservative:	21
Best Local Similarity:	53.30%	Mismatches:	38
Query Match:	56.05%	Indels:	40
DB:	10	Gaps:	5

US-09-701-674A-23 (1-179) x ADE25693 (1-2177)

```

QY      1 MetSerThrGlySerLeuSerAspValGluAspIeuGlnGluValGluMetLeuGluCys 20
      |||||
DQ      458 ATCTCCAGGGCGTCGGTCAGTGATCCGGAGAG-----ATGGAGCTT 499
      |||||

```

21 AspGlyLeuLys-----Me

500 CCGGGGTCACGGGAGTACCCGGTCCCGCCCTCCAGAGGCCGCCCTCCGGCGGTA 559

Qv 26 AspSerAsnLysGlu<sup>ph</sup>eValThrSerAsnGluSerThrGluGlu----- 40

Db 560 GAGCGCAGCTACGCTCGCCCACTGACAACTCGTCGGCAGAGGAGGAGGCCCGACGGC 619

[illegible]

D<sub>b</sub> 620 GAGGAGGAGCGCTGGCTCTGGGCACAGCCGGCAGCCGGGAGGCTGCAGAGGAGACGG 679

[illegible]

QY 49 ProGln---LysGlyArgGlyGlyLeuGlyLysArgArgLysAlaProThrLysLysSer 67  
 Db 680 CCCCGTGTGGCTGGGGCGGCGCGAGGTGGTAGCGGGCGGTGGTGGCAAGAGCCC 739  
 QY 68 ---ProLeuSerGlyValSerGlnGlyLysGlnValGlnArgAsnAlaAlaAla 86  
 Db 740 CTCGCCGCCAAGGCTCAGCGCAGAGTCAAGCAGTCCAGCGGAACGCGCCAACGCC 799  
 QY 87 ArgGluArgAlaArgMetArgValLeuSerLysAlaPheSerArgLeuLysThrThrLeu 106  
 Db 800 CGTAGCGTGGCGGATGCGGTGCTGAGCAAGCGCTTCCAGGCTCAAGACCGCTG 859  
 QY 107 ProTrpValProProAspThrLysLeuSerLysLeuAspThrLeuArgLeuAlaSerSer 126  
 Db 860 CCCTGGGTGCCCCGACACTAAGCTCTCCAGCTGGACACGCTCCGGCTGGCTTCCAGT 919  
 QY 127 TyrIleAlaHisLeuArgGlnIleLeuAlaAsnAspLysTyrGluAsnGlyTyrIleHis 146  
 Db 920 TACATCGCTCACCTGGCGGAGCTGTTGCAGGAGCGCTATGAGAACGGCTACGTGCAC 979  
 QY 147 ProValAsnLeuThrTrpProPheMetValAlaGlyLysProGluSerAspLeuLysGlu 166  
 Db 980 CCAGTGAACCTGACATGGCCATTCGTGGTCTCGGAGACCGGACTCTGACACCAAGAA 1039  
 QY 167 ValValThrAlaSerArgLeuCysGlyThrThrAla 178  
 Db 1040 GTTTCGCGAGCAACAGACTATGTGGAACCGCGCT 1075  
 RESULT 4  
 AAA72428  
 ID AAA72428 standard; cDNA; 2196 BP.  
 AC AAA72428;  
 DT 19-DEC-2000 (first entry)  
 XX Human nucleic acid-binding protein NuABP-47 cDNA.  
 XX Human nucleic acid-binding protein; NuABP; agonist; antagonist; EST;  
 KW expressed sequence tag; drug screening; recombinant expression; antibody;  
 KW reproductive disorder; infertility; immunological disorder;  
 KW neurological disorder; cell proliferative disorder; cancer; tumour; ss.  
 XX Homo sapiens.  
 OS  
 XX WO200044900-A2.  
 XX 03-AUG-2000.  
 XX 28-JAN-2000; 2000WO-US002237.  
 XX 29-JAN-1999; 99US-0117904P.  
 XX 29-JAN-1999; 99US-0117905P.  
 XX (INCY-) INCYTE PHARM INC.  
 XX Tang YT, Lal P, Hillman JL, Yue H, Azimzai Y, Lu AMD, Baughn MR;  
 XX Tran B, Shin LL, Au-Young JL;  
 XX WPI; 2000-499332/44.  
 XX P-PSDB; AAB21043.  
 XX Novel nucleic acid binding proteins, used to identify agonists and  
 XX antagonists of them, for the treatment of reproductive, immunological,  
 XX neurological and cell proliferative disorders including cancer.  
 XX Claim 4; Page 176; 180pp; English.  
 XX Sequences AAA72382-A72436 represent cDNAs encoding novel human nucleic  
 XX acid-binding proteins (NuABPs; AAB20997-B21051). These cDNAs were  
 XX produced by extension from an appropriate EST (expressed sequence tag)  
 XX using primers designed using the EST. The invention also relates to  
 XX expression constructs, host cells and transgenic organisms comprising a

CC human NuABP nucleic acid, recombinant production of the human NuABPs, and  
 CC antibodies against the human NuABPs, and also to methods of screening  
 CC modulators of human NuABP activity or expression. The human NuABPs, and  
 CC their agonists and antagonists are used to treat diseases associated with  
 CC overexpression or underexpression of functional NuABPs. Human NuABP  
 CC proteins and nucleotides, and NuABP agonists and antagonists can be used  
 CC to diagnose, treat and prevent reproductive, immunological, neurological  
 CC and cell proliferative disorders. Reproductive disorders that may be  
 CC treated using compositions of the invention include infertility,  
 CC endometriosis, disruptions of the menstrual cycle and disruptions of  
 CC spermatogenesis. Immunological disorders that may be treated include  
 CC AIDS, allergies, and autoimmune disorders such as multiple sclerosis,  
 CC rheumatoid arthritis, diabetes and systemic lupus erythematosus.  
 CC Neurodegenerative conditions such as Alzheimer's disease and Parkinson's  
 CC disease, prion diseases such as Creutzfeldt-Jakob disease, and mental  
 CC disorders such as schizophrenia. Cell proliferative disorders that may be  
 CC treated include a wide variety of cancers, and also arteriosclerosis,  
 CC atherosclerosis, cirrhosis and psoriasis  
 XX  
 SQ Sequence 2196 BP; 438 A; 691 C; 653 G; 414 T; 0 U; 0 Other;  
 Alignment Scores:  
 Pred. No.: 7 08e-47 Length: 2196  
 Score: 514.00 Matches: 113  
 Percent Similarity: 63.21% Conservatives: 21  
 Best Local Similarity: 53.30% Mismatches: 38  
 Query Match: 56.05% Indels: 40  
 DB: 3 Gaps: 5  
 US-09-701-674a-23 (1-179) x AAA72428 (1-2196)  
 QY 1 MetSerThrGlySerLeuSerAspValGluAspLeuGlnGluValGluMetLeuGluCys 20  
 Db 458 ATGTCCACGGCTCGGTGAGTGATCCGGAGGAG-----ATGGAGCTT 499  
 QY 21 AspGlyLeuLys-----Met 25  
 Db 500 CGGGGGTGCAGCGGGAGTAGTCCCGCTCCCAAGAGCGCCCTCCCGCGGTA 559  
 QY 26 AspSerAsnLysGluPheValThrSerAsnGluSerThrGluGlu-----40  
 Db 560 GAGCGCAGTACGCTCGCCAGTGCACACTCTCGCGCAGAGGAGGAGGACCCGACGGC 619  
 QY 41 -----SerSerAsnCysGluAsnGlySer 48  
 Db 620 GAGGAGGAGCGTCTGCTGGGCACAGCGCGCAGCGCGGAGCGTGCAGAGGAGGAGCGG 679  
 QY 49 ProGln---LysGlyArgGlyGlyLeuGlyLysArgArgLysAlaProThrLysLysSer 67  
 Db 680 CCCCGTGTGGCTGGGGCGGCGCGAGGTGGTAGCGGGCGGTGGTGGCAAGAGCCC 739  
 QY 68 ---ProLeuSerGlyValSerGlnGlyLysGlnValGlnArgAsnAlaAlaAla 86  
 Db 740 CTCGCCGCCAAGGCTCAGCGCAGAGTCAAGCAGTCCAGCGGAACGCGCCAACGCC 799  
 QY 87 ArgGluArgAlaArgMetArgValLeuSerLysAlaPheSerArgLeuLysThrThrLeu 106  
 Db 800 CGTAGCGTGGCGGATGCGGTGCTGAGCAAGCGCTTCCAGGCTCAAGACCGCTG 859  
 QY 107 ProTrpValProProAspThrLysLeuSerLysLeuAspThrLeuArgLeuAlaSerSer 126  
 Db 860 CCCTGGGTGCCCCGACACTAAGCTCTCCAGCTGGACACGCTCCGGCTGGCTTCCAGT 919  
 QY 127 TyrIleAlaHisLeuArgGlnIleLeuAlaAsnAspLysTyrGluAsnGlyTyrIleHis 146  
 Db 920 TACATCGCTCACCTGGCGGAGCTGTTGCAGGAGCGCTATGAGAACGGCTACGTGCAC 979  
 QY 147 ProValAsnLeuThrTrpProPheMetValAlaGlyLysProGluSerAspLeuLysGlu 166  
 Db 980 CCAGTGAACCTGACATGGCCATTCGTGGTCTCGGAGACCGGACTCTGACACCAAGAA 1039  
 QY 167 ValValThrAlaSerArgLeuCysGlyThrThrAla 178



XX Composition useful for diagnosis of conditions, disorders or diseases  
 PT associated with atherosclerosis, comprises several polynucleotides that  
 PT are differentially expressed in foam cell development.

XX Claim 1; Page 135-136; 315pp; English.

XX The present invention relates to the isolation of human polynucleotide  
 CC sequences that are differentially expressed during foam cell  
 CC differentiation. The polynucleotide sequences of the invention or a  
 CC composition comprising these polynucleotides are useful as a high  
 CC throughput method for detecting altered expression of one or more  
 CC polynucleotides in a sample. The polynucleotides can be used in the  
 CC diagnosis of disorders associated with foam cell development such as  
 CC atherosclerosis, cerebral stroke, and cardiovascular disorders such as  
 CC coronary artery disease. The polynucleotide sequences can also be used as  
 CC PCR primers and probes. The polynucleotides of the invention are also  
 CC useful in gene therapy. AAS94746-AAS95021 represent the human  
 CC polynucleotide sequences of the invention which are differentially  
 CC expressed during foam cell differentiation

XX Sequence 2382 BP; 459 A; 732 C; 702 G; 489 T; 0 U; 0 Other;

Alignment Scores:  
 Pred. No.: 4,77e-46 Length: 2382  
 Score: 507.00 Matches: 115  
 Percent Similarity: 63.13% Conservative: 22  
 Best Local Similarity: 53.00% Mismatches: 31  
 Query Match: 55.29% Indels: 49  
 DB: 6 Gaps: 6

US-09-701-674A-23 (1-179) x AAS94828 (1-2382)

Qy 1 MetSerThrGlySerLeuSerAspValGluAspLeuGlnGluValMetLeuGluCys 20  
 Db ATGTCACGGGCTCGGTGAGTATCGGAGGAG-----ATGGAGCTT 502  
 Qy 21 AspGlyLeuLys-----Met 25  
 Db CGGGGGCTGCAGCGGAGTACCGGTCCCGCCCTCCCAAGAGCGCGCCCTCCGCGCGGTA 562  
 Qy 26 AspSerAsnGlyGluPheValThrSerAsnGluSerThrGlu----- 39  
 Db GAGCGGAGTACCGCTCGCCAGTACGACCTCGTCGCGAGAGGAGGAGGAGCCCGCAGCGC 622  
 Qy 40 GluSerSerAsnGlyGluGlnGlySerProGlnGlyGlyArgGly----- 54  
 Db GAGGAGGAGCGCTCGCTCTGGGCACACAGCGCGGAGGAGGAGGAGGAGGAGGAGGAGG 682  
 Qy 55 -----GlyLeuGlyLysArgArgLys-----Ala 62  
 Db CCCCCGTGCTCGGGGCGGCGCGCGAGGTGAGTACCGCGCGGCGGTGGTGGCAAGAGGC 742  
 Qy 63 ProThrLysLysSerProLeuSerGlyValSerGln-GluGlyLysGlnValGlnArgAs 82  
 Db CCGCGCCCAAGGCTCAGCGGAGAGTGCAGACGACGACGACGACGACGACGACGACG 790  
 Qy 82 nAlaAlaAsnAlaArgGluArgAlaArgMetArgValLeuSerLysAlaPheSerArgLe 102  
 Db CGCGGCCCAAGCGCGTGGAGGTGCCCGATGCCGCTGAGCAAGAGCTTTCACAGGT 850  
 Qy 102 uLysThrThrLeuProTrpValProAspThrLysLeuSerLysLeuAspThrLeuAr 122  
 Db CAAGACCGAGCTCGCTCGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 910  
 Qy 122 gLeuAlaSerSerTyrIleAlaHisLeuArgGlnIleLeuAlaAsnAspLysTyrGluAs 142  
 Db GCTGGCTTCAGTATACCTCCTCAGTGGGAGCTGTTGACGAGGAGGAGGAGGAGGAG 970  
 Qy 142 nGlyTyrIleHisProValAsnLeuThrTrpProPheMetValAlaGlyLysProGluSe 162  
 Db CGGCTAGTGCACCCAGTGAACCTGACATGCCATTCTCGTGGTCTCGGGAAGACCGGACTC 1030

Qy 162 rAspLeuLysGluValValThrAlaSerArgLeuCysGlyThrThrAla 178  
 Db TGACACCAAGAAAGTTTCCGACGCAACAGACTATGTGGAACCAACCGCT 1079  
 RESULT 7  
 AAZ50464  
 ID AAZ50464 standard; cDNA; 1718 BP.  
 AC AAZ50464;  
 DT 23-MAY-2000 (first entry)  
 XX Murine muscudin bHLH protein encoding cDNA.  
 DE Muscudin; basic helix-loop-helix; bHLH; transcription factor; myogenesis;  
 KW screening; myogenic disease; aberration; muscle development; cytostatic;  
 KW gene therapy; diagnostic agent; muscular dystrophy; myopathy;  
 KW neuromuscular; skeletomuscular; myogenic cancer; mouse; ss.  
 OS Mus sp.  
 XX FH Key Location/Qualifiers  
 FT CDS 219..824  
 FT /\*tag= a  
 FT /product= "Murine muscudin bHLH protein"  
 FT /note= "Lacks Alanine-Threonine myogenic recognition motif"  
 XX WO200006720-A1.  
 XX 10-FEB-2000.  
 XX 30-JUL-1999; 99WO-AU000623.  
 XX 30-JUL-1998; 98AU-00004955.  
 XX (HALL-) HALL INST MEDICAL RES WALTER & ELIZA.  
 XX Robb L, Begley CG, Harvey RP;  
 DR WPI; 2000-195290/17.  
 DR P-PSDB; RAY44921.  
 PT New regulatory protein useful for gene therapy of myogenic cancer and for  
 PT detecting aberrations in muscle development in myogenic disease,  
 PT comprises protein associated with genes involved in myogenesis.  
 XX Claim 3; Fig 1A; 54pp; English.  
 CC The present cDNA sequence encodes murine muscudin bHLH (basic helix-loop-  
 CC helix) protein, which is a transcription factor associated with genes  
 CC involved in myogenesis. Muscudin gene can be used in genetic screening  
 CC for myogenic disease conditions, for detecting aberrations in muscle  
 CC development and in gene therapy. Anti-muscudin antibodies can be used as  
 CC therapeutic or diagnostic agents. Muscudin has cytostatic activity and  
 CC can be used to treat myogenic disease conditions like myopathies,  
 CC muscular dystrophies, neuromuscular and skeletomuscular disorders and  
 CC myogenic cancer  
 XX Sequence 1718 BP; 356 A; 477 C; 520 G; 365 T; 0 U; 0 Other;

Alignment Scores:  
 Pred. No.: 5.8e-45 Length: 1718  
 Score: 495.50 Matches: 112  
 Percent Similarity: 63.90% Conservative: 19  
 Best Local Similarity: 54.63% Mismatches: 43  
 Query Match: 54.03% Indels: 31  
 DB: 3 Gaps: 5

US-09-701-674A-23 (1-179) x AAZ50464 (1-1718)

Qy 1 MetSerThrGlySerLeuSerAspValGluAsp-----LeuGlnGluVal 15  
 |||||



ACH21857  
ID ACH21857 standard; cDNA; 482 BP.  
XX AC ACH21857;  
XX  
DT 13-OCT-2003 (first entry)  
XX  
XX Human adult ovary cDNA #237.  
XX  
XX Human; ss; sequencing by hybridisation; SBH; expressed sequence tag; EST;  
XX genome mapping; biodiversity; genetic disorder.  
XX  
XX Homo sapiens.  
XX  
XX US2003073623-A1.  
XX  
XX 17-APR-2003.  
XX  
XX 30-JUL-2001; 2001US-00918995.  
XX  
XX 30-JUL-2001; 2001US-00918995.  
XX  
XX (DRMA/) DRMANAC R T.  
XX (LABA/) LABAT I.  
XX (STAC/) STACHE-CRAIN B.  
XX (DICK/) DICKSON M C.  
XX (JONE/) JONES L W.  
XX  
XX Drmanac RT, Labat I, Stache-Crain B, Dickson MC, Jones LW;  
XX WPI; 2003-615964/58.  
XX  
XX New polynucleotide sequences obtained from various cDNA libraries, useful  
PT as hybridization probes, as oligomers for PCR, for chromosome and gene  
PT mapping, in the recombinant production of protein, or in generating  
PT antisense DNA or RNA.  
XX  
XX Claim 1; SEQ ID NO 9069; 44pp; English.  
XX  
XX The invention relates to an isolated polynucleotide comprising any one of  
CC 38043 cDNA sequences, appearing as ACH21857-ACH50831, whose sequence was  
CC determined by the technique of SBH (sequencing by hybridisation). Also  
CC included is a purified polypeptide comprising a sequence corresponding to  
CC a reading frame of the novel polynucleotide. The nucleic acid sequences  
CC are useful in diagnostics as expressed sequence tags (EST) for  
CC identifying expressed genes or for physical mapping of the human genome,  
CC in forensics, in assessing biodiversity, or in identifying mutations  
CC responsible for genetic disorders and other traits. The nucleotide  
CC sequences are also useful as hybridisation probes, as oligomers for PCR,  
CC for chromosome and gene mapping, in the recombinant production of  
CC protein, or in generating antisense DNA or RNA. The purified polypeptide  
CC is useful for generating antibodies specific for it. The present sequence  
CC is one of the 38043 isolated cDNA/EST sequences. Note: The sequence data  
CC for this patent did not form part of the printed specification, but was  
CC obtained in electronic format directly from USPTO at  
XX seqdata.uspto.gov/sequence.html?DocID=20030073623  
XX  
XX Sequence 482 BP; 99 A; 156 C; 143 G; 83 T; 0 U; 1 Other;

Alignment Scores:  
Pred. No.: 8,84e-45 Length: 482  
Score: 487.00 Matches: 95  
Percent Similarity: 98.96% Conservativity: 0  
Best Local Similarity: 98.96% Mismatches: 1  
Query Match: 53.11% Indels: 0  
DB: 9 Gaps: 0

US-09-701-674A-23 (1-179) x ACH21857 (1-482)

QY 84 AlaAsnAlaGluArgAlaArgMetArgValLeuSerLysAlaPheSerArgLeuLys 103  
DB 33 GCNCTAGCGGAGAGCGGCGCCGATCGAGTGCTGAGCAGGCCCTTCCAGACTCAAG 92

QY 104 ThrThrLeuProTrpValProAspThrLysLeuSerLysLeuAspThrLeuArgLeu 123  
DB 93 ACCACCCCTGGCTGGTGGCCCGACCAAGCTCTCCAAAGCTGGACACGCTCAGGCTG 152  
QY 124 AlaSerSerTrpIleAlaHisLeuArgGlnIleAlaAsnAspLysTrpGluAsnGly 143  
DB 153 GCGTCCAGCTACATCGCCCACTTGAGGCAGATCTCGCTAACGCAAAATACGGAACGGG 212  
QY 144 TyrIleHisProValAsnLeuThrTrpProPheMetValAlaGlyLysProGluSerAsp 163  
DB 213 TACATTCCACCGGTCAACCTGACGTGGCCCTTTATGGTGGCCGGAAACCCGAGAGTGAC 272  
QY 164 LeuLysGluValValThrAlaSerArgLeuCysGlyThrThrAlaSer 179  
DB 273 CTGAAGAGAGTGGTGACCGGAGCGGCTTATGTGGAACCAACCGCGTCC 320

RESULT 10  
ACA56430  
ID ACA56430 standard; cDNA; 240 BP.  
XX  
XX ACA56430;  
XX  
XX 06-JUN-2003 (first entry)  
XX  
XX Chicken signalling pathway polynucleotide probe SEQ ID NO 1028.  
XX  
XX Chicken; probe; ss; array element; Parkinson's disease;  
KW signalling pathway population; cancer; adenocarcinoma; leukaemia;  
KW immunopathy; AIDS; asthma; neuropathy; Alzheimer's disease; microarray.  
XX  
XX Gallus gallus.  
XX  
XX US6500938-B1.  
XX  
XX 31-DEC-2002.  
XX  
XX 30-JAN-1998; 98US-00016434.  
XX  
XX 30-JAN-1998; 98US-00016434.  
XX  
XX (INCY-) INCYTE GENOMICS INC.  
XX  
XX Au-Young J, Seilhamer JJ;  
XX WPI; 2003-352189/33.  
XX  
XX Combination of polynucleotide probes, useful as array elements in a  
PT microarray for monitoring the expression of a number of target  
PT polynucleotides.  
XX  
XX Claim 1; SEQ ID NO 1028; 65pp; English.  
XX  
XX The invention relates to a combination which, comprises a number of  
CC polynucleotide probes comprising a sequence selected from one of the 1490  
CC sequences mentioned in the specification. The combination is useful as an  
CC array element in a microarray for monitoring the expression of a number  
CC of target polynucleotides. The microarray is particularly useful in the  
CC diagnosis and treatment of cancer and immunopathology and neuropathology.  
CC The microarray is useful in diagnostics and treatment regimens, drug  
CC discovery and development, toxicological and carcinogenicity studies,  
CC forensics and pharmacogenomics. The microarray is also useful for  
CC monitoring progression of diseases and for developing sophisticated  
CC profiles for the effects of currently available therapeutic drugs. The  
CC combination is also useful for purifying a subpopulation of mRNAs, cDNAs  
CC and genomic fragments and in research and diagnostic applications. The  
CC array can detect changes in expression in a large number of genes coding  
CC for different signalling pathway populations which can be used to diagnose  
CC various diseases including cancer e.g. adenocarcinoma and leukaemia,  
CC immunopathies e.g. AIDS and asthma, neuropathies e.g. Alzheimer's disease  
CC and Parkinson's disease. The present sequence represents a polynucleotide  
CC probe of the invention. Note: The sequence data for this patent did not  
CC form part of the printed specification but was obtained in electronic  
CC format directly from USPTO at







XX PR 05-JUN-2000; 2000US-0209473P.  
 XX PR 05-JUN-2000; 2000US-0209531P.  
 XX PR 18-SEP-2000; 2000US-0233133P.  
 XX PR 18-SEP-2000; 2000US-0233617P.  
 XX PR 20-SEP-2000; 2000US-0234009P.  
 XX PR 20-SEP-2000; 2000US-0234034P.  
 XX PR 20-SEP-2000; 2000US-0234052P.  
 XX PR 22-SEP-2000; 2000US-0234509P.  
 XX PR 22-SEP-2000; 2000US-0234567P.  
 XX PR 25-SEP-2000; 2000US-0234923P.  
 XX PR 25-SEP-2000; 2000US-0234924P.  
 XX PR 25-SEP-2000; 2000US-0235077P.  
 XX PR 25-SEP-2000; 2000US-0235082P.  
 XX PR 25-SEP-2000; 2000US-0235134P.  
 XX PR 25-SEP-2000; 2000US-0235280P.  
 XX PR 26-SEP-2000; 2000US-0235637P.  
 XX PR 27-SEP-2000; 2000US-0235711P.  
 XX PR 27-SEP-2000; 2000US-0235720P.  
 XX PR 27-SEP-2000; 2000US-0235840P.  
 XX PR 27-SEP-2000; 2000US-0235863P.  
 XX PR 28-SEP-2000; 2000US-0236028P.  
 XX PR 28-SEP-2000; 2000US-0236032P.  
 XX PR 28-SEP-2000; 2000US-0236033P.  
 XX PR 28-SEP-2000; 2000US-0236034P.  
 XX PR 28-SEP-2000; 2000US-0236109P.  
 XX PR 28-SEP-2000; 2000US-0236111P.  
 XX PR 28-SEP-2000; 2000US-0236842P.  
 XX PR 29-SEP-2000; 2000US-0236891P.  
 XX PR 02-OCT-2000; 2000US-0237172P.  
 XX PR 02-OCT-2000; 2000US-0237173P.  
 XX PR 02-OCT-2000; 2000US-0237278P.  
 XX PR 02-OCT-2000; 2000US-0237294P.  
 XX PR 02-OCT-2000; 2000US-0237295P.  
 XX PR 02-OCT-2000; 2000US-0237316P.  
 XX PR 03-OCT-2000; 2000US-0237425P.  
 XX PR 03-OCT-2000; 2000US-0237598P.  
 XX PR 03-OCT-2000; 2000US-0237604P.  
 XX PR 03-OCT-2000; 2000US-0237606P.  
 XX PR 03-OCT-2000; 2000US-0237608P.  
 XX PR 01-NOV-2000; 2000US-0244867P.  
 XX PR 01-NOV-2000; 2000US-0245084P.  
 XX PA (AVAL-) AVALON PHARM.  
 XX PI Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;  
 XX PI Soppet DR, Weaver Z;  
 XX DR WPI; 2002-188264/24.  
 XX PT Screening for anti-neoplastic agent involves exposing cells to a chemical  
 XX PT agent to be tested for anti-neoplastic activity, and determining a change  
 XX PT in expression of a gene of a signature gene set.  
 XX PS Claim 1; SEQ ID NO 3856; 44pp; English.  
 XX PS The present invention describes a method (M1) for screening for an anti-  
 XX PS neoplastic agent. The method involves exposing cells to a chemical agent  
 XX PS to be tested for anti-neoplastic activity, determining a change in  
 XX PS expression of at least one gene (I) of a signature gene set, where (I)  
 XX PS comprises a sequence (S) selected from 847 sequences (given in ABL6164  
 XX PS to ABL70110), or is at least 95% identical to (S), where a change in  
 XX PS expression is indicative of anti-neoplastic activity. (I) has cytostatic  
 XX PS activity and can be used in gene therapy. M1 can be used for screening an  
 XX PS anti-neoplastic agent, and can be used for producing a product which is  
 XX PS the data collected with respect to the anti-neoplastic agent as a result  
 XX PS of M1, and the data is sufficient to convey the chemical structure and/or  
 XX PS properties of the agent. M1 can be used in the treatment of cancer such  
 XX PS as colon, breast, stomach, lung, thyroid, oesophageal, ovarian, kidney,  
 XX PS prostate or pancreatic cancer, adenocarcinoma, carcinoma, clear cell  
 XX PS cancer, infiltrating ductal cancer, infiltrating lobular cancer, squamous  
 XX PS cell carcinoma, neuroendocrine carcinoma, papillary carcinoma and Wilm's

CC tumour  
 XX Sequence 697 BP; 190 A; 164 C; 156 G; 180 T; 0 U; 7 Other;  
 SQ  
 Alignment Scores:  
 Pred. No.: 9.92e-26 Length: 697  
 Score: 318.00 Matches: 61  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 34.68% Indels: 0  
 DB: 6 Gaps: 0  
 US-09-701-674A-23 (1-179) x ABL65519 (1-697)  
 QY 119 AspThrLeuArgLeuAlaSerSerTyrIleAlaHisLeuArgGlnIleLeuAlaAsnApp 138  
 Db 3 GACACGCTCAGGTGGCGTCCAGCTACATCGCCCACTTGAGGCAGATCCTGTGTAACGAC 62  
 QY 139 LysTyrGluAsnGlyTyrIleHisProValAsnLeuThrTrpProPheMetValAlaGly 158  
 Db 63 AATACGAGAGCGGTACATCCCGTCACTGACGTGGCCCTTATGTGGACCAACCGCG 122  
 QY 159 LysProGluSerAspLeuLysGluValThrAlaSerArgLeuCysGlyThrAla 178  
 Db 123 AAACCCGAGAGTGACCTGAAAGAGTGGTGACCGCGCGCTTATGTGGACCAACCGCG 182  
 QY 179 Ser 179  
 Db 183 TCC 185  
 RESULT 13  
 ABL65935  
 ID ABL65935 standard; DNA; 697 BP.  
 XX AC ABL65935;  
 XX DT 15-MAY-2002 (first entry)  
 XX DE Lung cancer related gene sequence SEQ ID NO:4272.  
 XX KW Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;  
 KW stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;  
 KW cytostatic; gene therapy; antineoplastic; Wilm's tumour; adenocarcinoma;  
 KW gene; ds.  
 XX OS Homo sapiens.  
 XX PN WO200194629-A2.  
 XX PD 13-DEC-2001.  
 XX PF 30-MAY-2001; 2001WO-US010838.  
 XX PR 05-JUN-2000; 2000US-0209473P.  
 XX PR 05-JUN-2000; 2000US-0209531P.  
 XX PR 18-SEP-2000; 2000US-0233133P.  
 XX PR 18-SEP-2000; 2000US-0233617P.  
 XX PR 20-SEP-2000; 2000US-0234009P.  
 XX PR 20-SEP-2000; 2000US-0234034P.  
 XX PR 20-SEP-2000; 2000US-0234052P.  
 XX PR 22-SEP-2000; 2000US-0234509P.  
 XX PR 22-SEP-2000; 2000US-0234567P.  
 XX PR 25-SEP-2000; 2000US-0234923P.  
 XX PR 25-SEP-2000; 2000US-0234924P.  
 XX PR 25-SEP-2000; 2000US-0235077P.  
 XX PR 25-SEP-2000; 2000US-0235082P.  
 XX PR 25-SEP-2000; 2000US-0235134P.  
 XX PR 25-SEP-2000; 2000US-0235280P.  
 XX PR 26-SEP-2000; 2000US-0235637P.  
 XX PR 27-SEP-2000; 2000US-0235711P.  
 XX PR 27-SEP-2000; 2000US-0235720P.  
 XX PR 27-SEP-2000; 2000US-0235840P.

PR 27-SEP-2000; 2000US-0235863P.  
 PR 28-SEP-2000; 2000US-0236028P.  
 PR 28-SEP-2000; 2000US-0236032P.  
 PR 28-SEP-2000; 2000US-0236033P.  
 PR 28-SEP-2000; 2000US-0236034P.  
 PR 28-SEP-2000; 2000US-0236109P.  
 PR 28-SEP-2000; 2000US-0236111P.  
 PR 28-SEP-2000; 2000US-0236942P.  
 PR 29-SEP-2000; 2000US-0236891P.  
 PR 02-OCT-2000; 2000US-0237172P.  
 PR 02-OCT-2000; 2000US-0237278P.  
 PR 02-OCT-2000; 2000US-0237294P.  
 PR 02-OCT-2000; 2000US-0237295P.  
 PR 02-OCT-2000; 2000US-0237316P.  
 PR 03-OCT-2000; 2000US-0237425P.  
 PR 03-OCT-2000; 2000US-0237598P.  
 PR 03-OCT-2000; 2000US-0237604P.  
 PR 03-OCT-2000; 2000US-0237606P.  
 PR 03-OCT-2000; 2000US-0237608P.  
 PR 01-NOV-2000; 2000US-0244867P.  
 PR 01-NOV-2000; 2000US-0245084P.  
 XX  
 PA (AVAL-) AVALON PHARM.  
 XX  
 XX Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;  
 PI Soppet DR, Weaver Z;  
 XX  
 DR WPI; 2002-188264/24.  
 XX  
 XX Screening for anti-neoplastic agent involves exposing cells to a chemical  
 PT agent to be tested for anti-neoplastic activity, and determining a change  
 PT in expression of a gene of a signature gene set.  
 XX  
 XX Claim 1; SEQ ID NO 4272; 44pp; English.  
 PS  
 CC The present invention describes a method (M1) for screening for an anti-  
 CC neoplastic agent. The method involves exposing cells to a chemical agent  
 CC to be tested for anti-neoplastic activity, determining a change in  
 CC expression of at least one gene (I) of a signature gene set, where (I)  
 CC comprises a sequence (S) selected from 847 sequences (given in ABU61664  
 CC to ABU70110), or is at least 95% identical to (S), where a change in  
 CC expression is indicative of anti-neoplastic activity. (I) has cytostatic  
 CC activity and can be used in gene therapy. M1 can be used for screening an  
 CC anti-neoplastic agent, and can be used for producing a product which is  
 CC the data collected with respect to the anti-neoplastic agent as a result  
 CC of M1, and the data is sufficient to convey the chemical structure and/or  
 CC properties of the agent. M1 can be used in the treatment of cancer such  
 CC as colon, breast, stomach, lung, thyroid, oesophageal, ovarian, kidney,  
 CC prostate or pancreatic cancer, adenocarcinoma, carcinoma, clear cell  
 CC cancer, infiltrating ductal cancer, infiltrating lobular cancer, squamous  
 CC cell carcinoma, neuroendocrine carcinoma, papillary carcinoma and Wilms'  
 CC tumour  
 XX  
 SQ Sequence 697 BP; 190 A; 164 C; 156 G; 180 T; 0 U; 7 Other;  
 Alignment Scores:  
 Pred. No.: 9.92e-26 Length: 697  
 Score: 318.00 Matches: 61  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 34.68% Indels: 0  
 DB: 6 Gaps: 0  
 US-09-701-674A-23 (1-179) x ABL65935 (1-697)  
 QY 119 AspThrLeuArgLeuAlaSerSeryTyrIleAlaHisLeuArgGlnIleLeuAlaAsnAsp 138  
 DB 3 GACACGCTAGGCTGGCGTCACATCGCCCACTTGGGAGATCTCTGGCTAACGAC 62  
 QY 139 LysTyrGluAsnGlyTyrIleHisProValAsnLeuThrTrpProPheMetValAlaGly 158  
 DB 63 AAATACGAGAACGGGTACATTTCACCCGGTCAACCTGACGTGGCCCTTTATGTGGCGCGG 122

QY 159 LysProGluSerAspLeuLysGluValValThrAlaSerArgLeuGlyThrAla 178  
 DB 123 AAACCCGAGAGTGACCTGAAGAAGTGGTGACCGGAGCGCTTATGTGGACCGCG 182  
 QY 179 Ser 179  
 DB 183 TCC 185  
 RESULT 14  
 ABK64784  
 ID ABK64784 standard; DNA; 697 BP.  
 XX  
 AC ABK64784;  
 XX  
 DT 18-JUN-2002 (first entry)  
 XX  
 DE Human benign prostatic hyperplasia gene #679.  
 XX  
 KW Human; benign prostatic hyperplasia; BPH; prostate cancer; gene; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200212440-A2.  
 XX  
 PD 14-FEB-2002.  
 XX  
 PF 07-AUG-2001; 2001WO-US024708.  
 XX  
 PR 07-AUG-2000; 2000US-0223232P.  
 PR 05-JUN-2001; 2001US-00873319.  
 XX  
 PA (GENE-) GENE LOGIC INC.  
 PA (NISB ) JAPAN TOBACCO INC.  
 XX  
 PI Munger WE, Kulkarni P, Getzenberg RH, Waga I, Yamamoto J;  
 XX  
 WPI; 2002-257476/30.  
 XX  
 PT Identifying drugs for and diagnosing benign prostatic hyperplasia, by  
 PT detecting expression levels of one or more genes in prostate cells from  
 PT patient that are differentially regulated compared to normal prostate  
 XX cells.  
 PS Disclosure; Page 367; 44pp; English.  
 XX  
 CC The invention relates to a method of diagnosing (I) the onset or  
 CC progression of benign prostatic hyperplasia (BPH), or screening (II) for  
 CC or identifying an agent that modulates the onset or progression of BPH.  
 CC The method is based on changes in gene expression in BPH tissue isolated  
 CC from patients exhibiting different clinical states of prostate  
 CC hyperplasia as compared to normal prostate tissue. (I) comprises  
 CC detecting the expression levels of one or more genes in prostate cells  
 CC from the subject that are differentially regulated compared to normal  
 CC prostate cells. (II) comprises preparing a first gene expression profile  
 CC of BPH cells or BPH-like cell population, exposing the cells to the  
 CC agent, preparing a second gene expression profile of the agent exposed  
 CC cells, and comparing the first and second gene expression profiles. (I)  
 CC is useful for diagnosing the onset or progression of BPH. (II) is useful  
 CC for identifying an agent that modulates the onset or progression of BPH.  
 CC The methods are useful to present information identifying the expression  
 CC level in a tissue or cells, by comparing the expression level of genes  
 CC given in the specification in the tissue or cells to the level of  
 CC expression of gene in the database, and displaying the expression levels  
 CC of at least one gene in the tissue or cell sample compared to the  
 CC expression level in BPH. Agents using (II) are useful for treating BPH or  
 CC prostate cancer. ABK64106-ABK64860 represent human benign prostatic  
 CC hyperplasia gene sequences of the invention  
 XX  
 SQ Sequence 697 BP; 190 A; 164 C; 156 G; 180 T; 0 U; 7 Other;  
 Alignment Scores:  
 Pred. No.: 9.92e-26 Length: 697

Score: 318.00 Matches: 61  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 34.68% Indels: 0  
DB: 6 Gaps: 0

US-09-701-674A-23 (1-179) x ABK64784 (1-697)

QY 119 AsThrIeuAArgLeuAlaSerSerTyrIleAlaHisLeuAArgGlnIleLeuAlaSerMet 138  
DB 3 GACACGCTCAGGCTGGCGTCCAGCTACATCGCCCACTTCAGGCAGATCCTGGCTAACGAC 62  
QY 139 LysTyrGluAsnGlyTyrIleHisProValAsnLeuThrTrpProPheMetValAlaGly 158  
DB 63 AATACGAGACGGGTACATTACCCGGTCAACCTGACGTGGCCCTTTATGTGGCCGGG 122  
QY 159 LysProGluSerAspLeuIysGluValValThrAlaSerArgLeuCysGlyThrThrAla 178  
DB 123 AAACCCGAGAGTACCTGAAAGAGTGTGACCGCGAGCGCTTATGTGGAAACCCGCG 182  
QY 179 Ser 179  
DB 183 TCC 185

# RESULT 15

ACH23969  
ID ACH23969 standard; cDNA; 493 BP.

AC ACH23969;

DT 13-OCT-2003 (first entry)

DE Human adult ovary cDNA #2349.

Human; ss; sequencing by hybridisation; SSH; expressed sequence tag; EST; genome mapping; biodiversity; genetic disorder.

OS Homo sapiens.

XX US2003073623-A1.

XX 17-APR-2003.

XX 30-JUL-2001; 2001US-00918995.

XX 30-JUL-2001; 2001US-00918995.

XX (DRMA/) DRMANAC R T.

XX (LABA/) LABAT I.

XX (STAC/) STACHE-CRAIN B.

XX (DICK/) DICKSON M C.

XX (JONE/) JONES L W.

XX Drmanac RT, Labat I, Stache-Crain B, Dickson MC, Jones LW;

XX WPI; 2003-615964/59.

XX Claim 1; SEQ ID NO 11181; 44pp; English.

XX The invention relates to an isolated polynucleotide comprising any one of  
CC 38043 cDNA sequences, appearing as ACH2789-ACH50831, whose sequence was  
CC determined by the technique of SSH (sequencing by hybridisation). Also  
CC included is a purified polypeptide comprising a sequence corresponding to  
CC a reading frame of the novel polynucleotide. The nucleic acid sequences  
CC are useful in diagnostics as expressed sequence tags (EST) for  
CC identifying expressed genes or for physical mapping of the human genome,  
CC in forensic, in assessing biodiversity, or in identifying mutations  
CC responsible for genetic disorders and other traits. The nucleotide

CC sequences are also useful as hybridisation probes, as oligomers for PCR,  
CC for chromosome and gene mapping, in the recombinant production of  
CC protein, or in generating antisense DNA or RNA. The purified polypeptide  
CC is useful for generating antibodies specific for it. The present sequence  
CC is one of the 38043 isolated cDNA/EST sequences. Note: The sequence data  
CC for this patent did not form part of the printed specification, but was  
CC obtained in electronic format directly from USPTO at  
CC seqdata.uspto.gov/sequence.html?DocID=20030073623

XX Sequence 493 BP; 121 A; 136 C; 121 G; 112 T; 0 U; 3 Other;

## Alignment Scores:

Pred. No.: 6.54e-16 Length: 493  
Score: 228.00 Matches: 42  
Percent Similarity: 97.73% Conservative: 1  
Best Local Similarity: 95.45% Mismatches: 1  
Query Match: 24.86% Indels: 0  
DB: 9 Gaps: 0

US-09-701-674A-23 (1-179) x ACH23969 (1-493)

QY 136 AlaAsnAspLysTyrGluAsnGlyTyrIleHisProValAsnLeuThrTrpProPheMet 155  
DB 54 TCTAACGACAAATACGAGAACGGGTACATTCACCCGGTCAACCTGATGTGGCCCTTTATG 113  
QY 156 ValAlaGlyLysProGluSerAspLeuLysGluValValThrAlaSerArgLeuCysGly 175  
DB 114 GTGGCCGGAAACCCGAGAGTGACCTGAAAGAGTGTGACCGCGAGCGCTTATGTGGA 173  
QY 176 ThrThrAlaSer 179  
DB 174 ACCACCGCGTCC 185

Search completed: December 16, 2004, 22:12:34

Job time : 461 secs

Result No.	Score	Query		Length	DB	ID	Description
		Match	%				
1	1272	100.0	1272	6	BD211771	BD211771	Proteins
2	1252.4	98.5	1254	9	AF035718	AF035718	Homo sapi
3	1252.2	98.4	1257	9	AF047419	AF047419	Homo sapi
4	1229	96.6	1239	6	CQ718038	CQ718038	Sequence
5	809.4	63.6	3231	9	BC025697	BC025697	Homo sapi
6	776	61.0	1240	10	AF035717	AF035717	Mus muscu
7	774.4	60.9	1246	10	BC053525	BC053525	Mus muscu
8	764.4	60.1	1267	10	AF036945	AF036945	Mus muscu
9	761.6	59.9	1220	10	AB003453	AB003453	Mus muscu
10	759.6	59.7	1217	10	AF029753	AF029753	Mus muscu
11	752	59.1	1202	10	AF047418	AF047418	Mus muscu
12	717.2	56.4	25186	9	AL356109	AL356109	Human DNA
13	581	45.7	697	6	AX333347	AX333347	Sequence
14	581	45.7	697	6	AX333763	AX333763	Sequence
15	537	42.2	537	9	CR450293	CR450293	Homo sapi
16	466.2	36.7	471	6	AX333399	AX333399	Sequence
17	466.2	36.7	471	6	AX333949	AX333949	Sequence
18	424.6	33.4	187332	2	AC101292	AC101292	Mus muscu
19	406.2	31.9	487	9	HSPAI5B	Z78720 H. sapiens f	

Query Match	100.0%;	Score 1272;	DB 6;	Length 1272;	
Best Local Similarity	100.0%;	Pred. No. 7e-249;			
Matches 1272;	Conservative	0;	Mismatches	0;	Indels
Qy	1	TCACGGCCACGACTCTGGAGTGGGAAACAGAGACCGGTTCTCTGCTGCAGAGTC	60		
Db	1	TCACGGCCACGACTCTGGAGTGGGAAACAGAGACCGGTTCTCTGCTGCAGAGTC	60		
Qy	61	CTCGGGTTCCTTCTCACAACTCTCGAAGGGGAAAGGTTGTGAGACCCACAGACCC	120		
Db	61	CTCGGGTTCCTTCTCACAACTCTCGAAGGGGAAAGGTTGTGAGACCCACAGACCC	120		
Qy	121	CAACTCAGTCCAGACGAGGTGCTGCGCCACACTCGGAGGCTCTTGGTTTCAGG	180		
Db	121	CAACTCAGTCCAGACGAGGTGCTGCGCCACACTCGGAGGCTCTTGGTTTCAGG	180		
Qy	181	GTCTCTGTCTCTCTCTCACCCCTCTTCTCGCTTCTCTGTCTCTCTCTCTCT	240		
Db	181	GTCTCTGTCTCTCTCTCACCCCTCTTCTCGCTTCTCTGTCTCTCTCTCTCT	240		
Qy	241	CTCTCTCTCTCTCTCTCTCCCAACATGTCACCGGCTCCCTCAGCGATGTGAGAC	300		
Db	241	CTCTCTCTCTCTCTCTCTCCCAACATGTCACCGGCTCCCTCAGCGATGTGAGAC	300		
Qy	301	CTTCAGAGGTGAGATGTTGGAATGTGACGGGTTGAAATGATTCGAAACAGAAATTT	360		
Db	301	CTTCAGAGGTGAGATGTTGGAATGTGACGGGTTGAAATGATTCGAAACAGAAATTT	360		
Qy	361	GTGACTTCCAAACAGAGACCCGAGGAGAGCTCACTCGGAGATGGTCTCTCCAGAG	420		
Db	361	GTGACTTCCAAACAGAGACCCGAGGAGAGCTCACTCGGAGATGGTCTCTCCAGAG	420		
Qy	421	GGCGGGGGCTGGGCAAGAGAGAGAGGGCCGCCAACAGAGAGCCCTCTGAGCGG	480		
Db	421	GGCGGGGGCTGGGCAAGAGAGAGAGGGCCGCCAACAGAGAGCCCTCTGAGCGG	480		
Qy	481	GTGAGCAGAGAGGGGAGGAGTCCAGCGCAACCGCGCAACCGCGAGAGCGGCGCG	540		
Db	481	GTGAGCAGAGAGGGGAGGAGTCCAGCGCAACCGCGCAACCGCGAGAGCGGCGCG	540		
Qy	541	ATCGAGTGTGACAGAGGCTTCTCAGACTCAGACACACCTGCGTGGTGGCGCC	600		
Db	541	ATCGAGTGTGACAGAGGCTTCTCAGACTCAGACACACCTGCGTGGTGGCGCC	600		
Qy	601	GACACCAAGTCTCCAAAGTGCAGAGCTCAGCGTGGCGTCCAGGTACATCGCCACTTG	660		
Db	601	GACACCAAGTCTCCAAAGTGCAGAGCTCAGCGTGGCGTCCAGGTACATCGCCACTTG	660		
Qy	661	AGGAGATCTGCTTAACGACAAATACGAGACGGGTACATTCACCGCGTCACTGACG	720		
Db	661	AGGAGATCTGCTTAACGACAAATACGAGACGGGTACATTCACCGCGTCACTGACG	720		
Qy	721	TGCGCTTTATGTGCGCGGAAACCCGAGAGTGACTCAAGAAAGTGTGACCGCGAGC	780		
Db	721	TGCGCTTTATGTGCGCGGAAACCCGAGAGTGACTCAAGAAAGTGTGACCGCGAGC	780		
Qy	781	CGCTTATGTGGAACACCGCGTCTGACCTTGGAGTGGAGTCTGGGAAAGGCGCGTC	840		
Db	781	CGCTTATGTGGAACACCGCGTCTGACCTTGGAGTGGAGTCTGGGAAAGGCGCGTC	840		
Qy	841	CCGGGGGAGCGGGCCCCGGGAGGGGACCCCTGCGCTCAGTGTCTCTGCTCTGCTTC	900		
Db	841	CCGGGGGAGCGGGCCCCGGGAGGGGACCCCTGCGCTCAGTGTCTCTGCTCTGCTTC	900		
Qy	901	CCCTCTGCAATGCTCTCTCTCTGTCGCCACCCCGGAGAACACTTTACAGACGAGGAG	960		
Db	901	CCCTCTGCAATGCTCTCTCTCTGTCGCCACCCCGGAGAACACTTTACAGACGAGGAG	960		
Qy	961	ATTGCTTCCAAACAGAGAGATCAATGTACTTACAGAGATTCCTATCTATTTAACTT	1020		
Db	961	ATTGCTTCCAAACAGAGAGATCAATGTACTTACAGAGATTCCTATCTATTTAACTT	1020		
Qy	1021	TATTAACCTTACCGGTGAATGACTCTGCAAGCCCTGCTGCTCCAAAGTCAATATTAAT	1080		

Db	1021	TATTAACCTTACCGGTGAATGACTCTGCAAGCCCTGCTGCTCCAAAGTCAATATTAAT	1080		
Qy	1081	ATAAATATATAATAATAGATAAGAGCCATCAATGTATCTTTTGTACAATATGTTGTA	1140		
Db	1081	ATAAATATATAATAATAGATAAGAGCCATCAATGTATCTTTTGTACAATATGTTGTA	1140		
Qy	1141	GTAGATCATAGATAGCTGACTTTGACAGTCAATTTATAAGTAATTCATTAAGATA	1200		
Db	1141	GTAGATCATAGATAGCTGACTTTGACAGTCAATTTATAAGTAATTCATTAAGATA	1200		
Qy	1201	TATATTTTTTCAAAACAAGTTTTGCTACTTTTGAATAAATACTTTCTTATATGCTAA	1260		
Db	1201	TATATTTTTTCAAAACAAGTTTTGCTACTTTTGAATAAATACTTTCTTATATGCTAA	1260		
Qy	1261	AAAAAATAAATAA 1272			
Db	1261	AAAAAATAAATAA 1272			
RESULT 2					
LOCUS	AF035718	1254 bp	mRNA	linear	PRI 05-OCT-1998
DEFINITION	Homo sapiens mesoderm-specific basic-helix-loop-helix protein				
ACCESSION	AF035718				
VERSION	AF035718				
KEYWORDS	AF035718.1 GI:2745886				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
REFERENCE	1 (bases 1 to 1254)				
AUTHORS	Quaggin, S.E., Vanden Heuvel, G.B. and Igarashi, P.				
TITLE	Pod-1, a mesoderm-specific basic-helix-loop-helix protein expressed in mesenchymal and glomerular epithelial cells in the developing kidney				
JOURNAL	Mech. Dev. 71 (1-2), 37-48 (1998)				
MEDLINE	98175875				
PUBMED	9507058				
REFERENCE	2 (bases 1 to 1254)				
AUTHORS	Quaggin, S.E., Vanden Heuvel, G.B. and Igarashi, P.				
TITLE	Direct Submission				
JOURNAL	Submitted (24-NOV-1997) Internal Medicine, Yale University, 333 Cedar Street, New Haven, CT 06520-8029, USA				
FEATURES	Location/Qualifiers				
source	1..1254				

QY 8 CCACGACTCTGGAGTGGGGAACACAGAGACCGGTTCTCTCTGTCGAGAAAGTCTCTCGGG 67  
 Db 1 CCACGACTCTGGAGTGGGGAACACAGAGACCGGTTCTCTCTGTCGAGAAAGTCTCTCGGG 60  
 QY 68 TTCCTTCTCACAACCTCTGGAGGGGAAAGGTTGTGAGACCCACACAGACCCCACTCC 127  
 Db 61 TTCCTTCTCACAACCTCTGGAGGGGAAAGGTTGTGAGACCCACACAGACCCCACTCC 120  
 QY 128 AGCTCCACGAGAGTGGTGGCCACACACTCGGAGGGCTCTTGTTTCAGGGTCTCTC 187  
 Db 121 AGCTCCACGAGAGTGGTGGCCACACACTCGGAGGGCTCTTGTTTCAGGGTCTCTC 180  
 QY 188 TGCTCTCTCTCACCCTC 247  
 Db 181 TGCTCTCTCTCACCCTC 240  
 QY 248 CCTCGTCCACTCCCCCAACATGTCACCGGTTCCCTCAGCATGTGAGAGACCTTCAAG 307  
 Db 241 CCTCGTCCACTCCCCCAACATGTCACCGGTTCCCTCAGCATGTGAGAGACCTTCAAG 300  
 QY 308 AGTGGAGATGTTGGAATGTGAGGGTGTGAATGATTGGAATGGAATGGAATGGAATGGAAT 367  
 Db 301 AGTGGAGATGTTGGAATGTGAGGGTGTGAATGATTGGAATGGAATGGAATGGAATGGAAT 360  
 QY 368 CCAACGAGACCGGAGAGTCCCACTCGAGATGGTCTCTCCAGAGAGGGCCG 427  
 Db 361 CCAACGAGACCGGAGAGTCCCACTCGAGATGGTCTCTCCAGAGAGGGCCG 420  
 QY 428 GCGGCTGGGCAAGAGAGGAGGAGGCGCCACCAAGAGAGGCGGCTGAGCGGGTCAAGC 487  
 Db 421 GCGGCTGGGCAAGAGAGGAGGAGGCGCCACCAAGAGAGGCGGCTGAGCGGGTCAAGC 480  
 QY 488 AGGAGGGAGAGGTCGAGCAACCGCGCAACCGCGAGAGGCGGCGCGCATGCGAG 547  
 Db 481 AGGAGGGAGAGGTCGAGCAACCGCGCAACCGCGAGAGGCGGCGCGCATGCGAG 540  
 QY 548 TGCTGACGAGGCTTCTCCAGACTCAAGACCACTGCGGTCGCGGTCGCGGTCGCGGTCGCGG 607  
 Db 541 TGCTGACGAGGCTTCTCCAGACTCAAGACCACTGCGGTCGCGGTCGCGGTCGCGGTCGCGG 600  
 QY 608 AGCTCTCAAGCTGGAACGCTGAGGTCGAGTCCAGTACATCGCCCACTTGAAGCAGA 667  
 Db 601 AGCTCTCAAGCTGGAACGCTGAGGTCGAGTCCAGTACATCGCCCACTTGAAGCAGA 660  
 QY 668 TCCTGGCTAACGACAATACGAGACGGGTACATTCACCGGTCACCTGAGTGGCCCT 727  
 Db 661 TCCTGGCTAACGACAATACGAGACGGGTACATTCACCGGTCACCTGAGTGGCCCT 720  
 QY 728 TTATGTGGCCGGAAACCGAGAGTGAACCTGAAGAAGTGTGACCGGAGCCGCTTAT 787  
 Db 721 TTATGTGGCCGGAAACCGAGAGTGAACCTGAAGAAGTGTGACCGGAGCCGCTTAT 780  
 QY 788 GTGGAACCAACCGGTCCTGACCTTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAG 847  
 Db 781 GTGGAACCAACCGGTCCTGACCTTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAG 840  
 QY 848 GAGCGGCGCCCGGAGAGGAGCCCTCGCCCTCAGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 907  
 Db 841 GAGCGGCGCCCGGAGAGGAGCCCTCGCCCTCAGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 900  
 QY 908 CAATGTCT 967  
 Db 901 CAATGTCT 960  
 QY 968 TCCAAACCAAGAGAGATCAATTTGACTTACAAAGATTCCTCATCTATTAACTTTATTAC 1027  
 Db 961 TCCAAACCAAGAGAGATCAATTTGACTTACAAAGATTCCTCATCTATTAACTTTATTAC 1020  
 QY 1028 TTCTACCGTGAATGATCTGCAAGCCTTGCTGGTCCAGTGCATGTAATGTAATTAATAAATA 1087  
 Db 1021 TTCTACCGTGAATGATCTGCAAGCCTTGCTGGTCCAGTGCATGTAATGTAATTAATAAATA 1080

QY 1088 TATAAATAGATAAGAGCGCTATCAATGATCTTTTGTACATATATGTTGTAAATGTAGATC 1147  
 Db 1081 TATAAATAGATAAGAGCGCTATCAATGATCTTTTGTACATATATGTTGTAAATGTAGATC 1140  
 QY 1148 ATAGATAGCTGACTTTTGACAGTCCATTTATAAGTAATTCACCTTAAAGATATATTT 1207  
 Db 1141 ATAGATAGCTGACTTTTGACAGTCCATTTATAAGTAATTCACCTTAAAGATATATTT 1200  
 QY 1208 TTTTCAACAAGTTTGTCTACTTTTGAATAAATAATCTTTCTTTATATGCTAAA 1261  
 Db 1201 TTTTCAACAAGTTTGTCTACTTTTGAATAAATAATCTTTCTTTATATGCTAAA 1254

RESULT 3  
 AF047419 1257 bp mRNA linear PRI 06-OCT-1998  
 LOCUS Homo sapiens epicardin mRNA, complete cds.  
 DEFINITION  
 ACCESSION AF047419  
 VERSION AF047419.1 GI:2921852  
 KEYWORDS  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 1257)  
 AUTHORS Robb,L., Mifsud,L., Hartley,L., Biben,C., Copeland,N.G., Gilbert,D.J., Jenkins,N.A. and Harvey,R.P.  
 TITLE epicardin: A novel basic helix-loop-helix transcription factor gene of developing lung, gut, kidney, and gonads  
 JOURNAL Dev. Dyn. 213 (1), 105-113 (1998)  
 MEDLINE 98401941  
 PUBMED 9733105  
 REFERENCE 2 (bases 1 to 1257)  
 AUTHORS Robb,L. and Mifsud,L.  
 TITLE Direct Submission  
 JOURNAL Submitted (10-FEB-1998) Haematology and Cancer, Walter and Eliza Hall Institute of Medical Research, PO Royal Melbourne Hospital, Melbourne, VIC 3050, Australia  
 FEATURES  
 Location/Qualifiers  
 1..1257  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 260..799  
 /function="basic helix-loop-helix transcription factor"  
 /note="expressed in embryonic epicardium and mesenchyme-derived tissues of lung, gut, kidney, and gonad"  
 /codon\_start=1  
 /product="epicardin"  
 /protein\_id="AAC62532.1"  
 /db\_xref="GI:2921853"  
 /translation="MSTGSLSDVEDLQEVLEKCDGLKMDSNKEFVTSNBSSTSSNC  
 ENSPQKRGRLGKRRKAPTKPSLGSVQEGKQVQRTANARERAMRVLSKAFPSRL  
 KTLTPWPPPTKLKLDLTLASSYIAHLRQILANDKIYENGYIHPVNLITWFFWVAGRP  
 ESDLKEVVTVAIRLCGTTAS"

Query Match 98.4%; Score 1252.2; DB 9; Length 1257;  
 Best Local Similarity 99.8%; Pred. No. 7.6e-245;  
 Matches 1254; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 9 CACGACTCTGGAGTGGGAAACAGAGAGCGGTTCTCTCTGTCGAGAGTCTCTCGGGGT 68  
 Db 1 CACGACTCTGGAGTGGGAAACAGAGAGCGGTTCTCTCTGTCGAGAGTCTCTCGGGGT 60  
 QY 69 TCCTTCTCACAACCTCTCGAGAGGGAAGGTTGTGAGACCCACACAGACCCCACTCCA 128  
 Db 61 TCCTTCTCACAACCTCTCGAGAGGGAAGGTTGTGAGACCCACACAGACCCCACTCCA 120  
 QY 129 GCTCCACGAGAGGAGTGGCTGCGCCACACTCGGAGGCGCTCTTGGTTTCAGGGTCTCTCT 188

*Sept, bad date*



```

Db 601 ACCAAGCTCTCCAAAGCTGGACACGCTCAGCTGCGGCTCCAGCTACATCCGCCCACTTGAGG 660
Qy 664 CAGATCCTGCTGCTAACGACAAATACGAGAAACGGGTACATTCACCCCGCTCAACCTGACGTGG 723
Db 661 CAGATCCTGCTGCTAACGACAAATACGAGAAACGGGTACATTCACCCCGCTCAACCTGACGTGG 720
Qy 724 CCCTTTATGTTGGTGGCGGGAACCCGAGAGTGCCTGAAGAAGTGTGTGACCCGCGAGCCGC 783
Db 721 CCCTTTATGTTGGTGGCGGGAACCCGAGAGTGCCTGAAGAAGTGTGTGACCCGCGAGCCGC 780
Qy 784 TTAATGGAACACACCCGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 843
Db 781 TTAATGGAACACACCCGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 840
Qy 844 GGGGAGCGGGGCCCCGGGAAGGCGACCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 903
Db 841 GGGGAGCGGGGCCCCGGGAAGGCGACCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 900
Qy 904 CTGCGAATGCTCTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 963
Db 901 CTGCGAATGCTCTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 960
Qy 964 CGTTTCCAAACACGAGAGATCAATTTGATCTTACAAAGATTCCTCACTATTTAACTTTAT 1023
Db 961 CGTTTCCAAACACGAGAGATCAATTTGATCTTACAAAGATTCCTCACTATTTAACTTTAT 1020
Qy 1024 TACTTCTACCGTGAATGCTGCAAGCTTGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1083
Db 1021 TAACTTCTACCGTGAATGCTGCAAGCTTGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1080
Qy 1084 AATATATAATAGATAAGAGCCCTATCAATGTATCTTTTGTACAAATATGTTGAAAAATGTA 1143
Db 1081 AATATATAATAGATAAGAGCCCTATCAATGTATCTTTTGTACAAATATGTTGAAAAATGTA 1140
Qy 1144 GATCATAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1203
Db 1141 GATCATAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1200
Qy 1204 ATTTTTCACAAAGTTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1262
Db 1201 ATTTTTCACAAAGTTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1259

BC025697 3231 bp mRNA linear PRI 30-JUN-2004
Homo sapiens transcription factor 21, transcript variant 1, mRNA
LOCUS BC025697
DEFINITION BC025697.1 GI:19344015
ACCESSION BC025697
VERSION BC025697.1
KEYWORDS MGC.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 3231)
Strausberg,R.L., Feilungold,E.A., Grouse,L.H., Derge,J.G.,
Klausner,R.D., Collins,F.S., Wagner,L., Shemen,C.M., Schuler,G.D.,
Altschul,S.F., Zebner,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
Scapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
Villalón,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
Fahey,J., Helton,E., Kettelman,M., Madan,A., Rodriguez,S.,
Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smallos,D.E.,
Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
12477932
2 (bases 1 to 3231)
Strausberg,R.
Direct Submission
Submitted (06-MAR-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: cgapbs-remail.nih.gov
Tissue Procurement: Life Technologies, Inc.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: National Institutes of Health Intramural
Sequencing Center (NISC),
Gaithersburg, Maryland;
Web site: http://www.nisc.nih.gov/
Contact: nisc.mgc@nhgri.nih.gov
Akhter,N., Ayele,K., Beckstrom-Sternberg,S.M., Benjamin,B.,
Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C., Brooks,S.,
Dietrich,N.L., Granite,S., Guan,X., Gupta,J., Haghighi,P.,
Hansen,N.L., Ho,S.-L., Karlins,E., Kwong,P., Laric,P., Legaspi,R.,
Maduro,Q.L., Mabilio,C., Maskeri,B., Mastrian,S.D., McCloskey,J.C.,
McDowell,J., Pearson,R., Stantripop,S., Thomas,P.J., Touchman,J.W.,
Turgeon,C., Vogt,J.L., Walker,M.A., Wetherby,K.D., Wiggins,L.,
Young,A., Zhang,L.-H. and Green,E.D.
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAC Plate: 49 Row: K Column: 7
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 4507394.
FEATURES
Location/Qualifiers
1..3231
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="MGC:34534 IMAGE:5221575"
/tissue_type="Pancreas, Spleen, adult pooled"
/clone_id="NIH_MGC_120"
/lab_host="DH10B"
/notes="Vector: pCMV-SPORT6"
1..3231
/gene="TCF21"
/notes="synonym: POD1"
/db_xref="LocusID:6943"
/db_xref="MIM:603306"
260..799
/gene="TCF21"
/codon_start=1
/product="transcription factor 21"
/protein_id="AAH25697.1"
/db_xref="GI:19344016"
/db_xref="LocusID:6943"
/db_xref="MIM:603306"
/translation="MSTGSLSDYEDLQEVLECDGLKMDSNKEFVTSNBSTSSNC
ENGSPQKRGKGLGRRKAPTKSKPLSCGVSGEQVQRNANARERAMRVLSKAFSRLL
KTLTPWPPDPTKSLKDLTKLASSYIAHLQIILANDKYENGYIHPVNLTPFWVWAGKP
ESDLKEVTVTASRLCGTTAS"
ORIGIN
Query Match 63.6%; Score 809.4; DB 9; Length 3231;
Best Local Similarity 98.1%; Pred. No. 1.6e-154;
Matches 819; Conservative 0; Mismatches 16; Indels 0; Gaps 0;
Qy 9 CACGACTCTGGAGTGGGAAACAGAGACCGGTTCTCTCTGCTCAGAGTCTCGGGGT 68
Db 1 CACGACTCTGGAGTGGGAAACAGAGACCGGTTCTCTCTGCTCAGAGTCTCGGGGT 60

```



```

QY 69 TCCTTCTCACAACCTCTGCGAAGGGGAAAGGTTGTGAGACCCCAACAGACCCCAACTCCA 128
Db 61 TCCTTCTCACAACCTCTGCGAAGGGGAAAGGTTGTGAGACCCCAACAGACCCCAACTCCA 120
QY 129 GCTCCAGCAGGAGGTTGCGCCACATCTCGGAGGCTCTTGTTTCAGGGTCTCTCT 188
Db 121 GCTCCAGCAGGAGGTTGCGCCACATCTCGGAGGCTCTTGTTTCAGGGTCTCTCT 180
QY 189 GTCTCTCTCACCTCTTCTCGCTTTCTCTGTCTCTCTGTCTCTCTCTCTCTCTCTCC 248
Db 181 GTCTCTCTCACCTCTTCTCGCTTTCTCTGTCTCTCTGTCTCTCTCTCTCTCTCTCC 240
QY 249 CTGCTCACTCCCAACATGTCACCGCTCCCTCAGCGATGTGAGAGCCTTCAAGA 308
Db 241 CTGCTCACTCCCAACATGTCACCGCTCCCTCAGCGATGTGAGAGCCTTCAAGA 300
QY 309 GGTGGAGATGTTGGAATGTGAAGGTTGAAATGGATTGAAACAGGAATTTGTGACTTC 368
Db 301 GGTGGAGATGTTGGAATGTGACGGTTGAAATGGATTGAAACAGGAATTTGTGACTTC 360
QY 369 CACGAGAGCACCGAGGAGCTCCAACTGCGGATGAGTCTCCCAAGAGGCGCGGG 428
Db 361 CACGAGAGCACCGAGGAGCTCCAACTGCGGATGAGTCTCCCAAGAGGCGCGGG 420
QY 429 CGGCCTGGGCAAGAGGAGGAGCGCCACCAAGAGAGCCCTTGAAGCGGGTCAAGCA 488
Db 421 CGGCCTGGGCAAGAGGAGGAGCGCCACCAAGAGAGCCCTTGAAGCGGGTCAAGCA 480
QY 489 GAGGGGAGCAGGTTCAGCGCAACCGCGCAACCGCGAGAGCGGCGCGCATGCGAGT 548
Db 481 GAGGGGAGCAGGTTCAGCGCAACCGCGCAACCGCGAGAGCGGCGCGCATGCGAGT 540
QY 549 GCTGAGCAAGGCTTCTCCAGACTCAAGACCACTGCGCTGCGTCCCGCCGACACAA 608
Db 541 GCTGAGCAAGGCTTCTCCAGACTCAAGACCACTGCGCTGCGTCCCGCCGACACAA 600
QY 609 GCTCTCAAGCTGGACACGCTAGGCTGGGTCAGCTACATCGGCCCATCTGAGGAGAT 668
Db 601 GCTCTCAAGCTGGACACGCTAGGCTGGGTCAGCTACATCGGCCCATCTGAGGAGAT 660
QY 669 CTGGCTTAACGACAAATACGAGAACGGGTACATTACCCGGTCAACTGACGTGCGCTT 728
Db 661 CTGGCTTAACGACAAATACGAGAACGGGTACATTACCCGGTCAACTGACGTGCGCTT 720
QY 729 TATGGTGGCGGAAACCCGAGAGTCACTGAAAGAGTGGTGAACGCGAGCGGCTTATG 788
Db 721 TATGGTGGCGGAAACCCGAGAGTCACTGAAAGAGTGGTGAACGCGAGCGGCTTATG 780
QY 789 TGAACCAACCGCTGCTGACCTTGGAGTGGGTTGGTACCTGGAGCGAGCTTCTCCG 843
Db 781 TGAACCAACCGCTGCTGACCTTGGAGTGGGTTGGTACCTGGAGCGAGCTTCTCCG 835

RESULT 6
AF035717 1240 bp mRNA linear ROD 05-OCT-1998
LOCUS AF035717
DEFINITION Mus musculus mesoderm-specific basic-helix-loop-helix protein
ACCESSION AF035717
VERSION AF035717
KEYWORDS (Pod1) mRNA, complete cds.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
TITLE 1 (bases 1 to 1240)
REFERENCE Quaggin, S.E., Vanden Heuvel, G.B. and Igarashi, P.
AUTHORS Pod-1, a mesoderm-specific basic-helix-loop-helix protein expressed
TITLE in mesenchymal and glomerular epithelial cells in the developing
JOURNAL kidney. Dev. 71 (1-2), 37-48 (1998)
MEDLINE 98175875

```

```

PUBMED 9507058
REFERENCE 2 (bases 1 to 1240)
AUTHORS Quaggin, S.E., Vanden Heuvel, G.B. and Igarashi, P.
TITLE Direct Submission
JOURNAL Submitted (24-NOV-1997) Internal Medicine, Yale University, 333
Cedar Street, New Haven, CT 06520-8029, USA
FEATURES
source
1. 1240
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/chromosome="10"
/map="near D10Mit2"
2. 1240
/gene="Pod1"
226..765
/gene="Pod1"
/notes="Pod-1"
/codon_start=1
/product="mesoderm-specific basic-helix-loop-helix
protein"
/db_xref="GI:2745885"
/translation="MSGSLSDVEDLQEVEMLDCLSKVDSNKEFGTSTNESTEEGSGNC
ENGSPKRGGLGKRRKAPTKSPGVSQEGKQVORNAARARMRVLSKAFRL
KTLLPWPDPDKSLDLTLRLASSYIAHLRQILANDKYENGYIHPVNLTPFWVAGKP
ENDLKEVVYANRLCGTTAS"
457..618
/misc_feature
/gene="Pod1"
/notes="encodes basic-helix-loop-helix domain"

ORIGIN
Query Match 61.0%; Score 776; DB 10; Length 1240;
Best Local Similarity 79.4%; Pred. No. 8.3e-148;
Matches 1020; Conservative 0; Mismatches 200; Indels 64; Gaps 6;

QY 3 TAGGGCAGCACTCTGGAGTGGGAAACAGAGAGCGGTTCTCTGCTGCAGAGTCCCT 62
Db 6 TAAGGCAGCACTCCAGTGGTGGGAGATAGACATGGTTTCCCTGTGCGGGAATCCCT 65
QY 63 CGGGGTTCTTCTCACAACCTCTCGAAGGGGAAAGGTTGTGAGACCCCAACAGACCCCA 122
Db 66 TGTGATCGTTTCCAGCACTCTGAAAGGAAAGCTGGGGAGCGGACTGCCAGATCCCA 125
QY 123 ACTCAGCTCCAGCAGGAGTGGCTGCGCCACATCTCGGAGGCTCTTGGTTTCAGGTT 182
Db 126 CTTCAACCCCAACACAGAAAGTGGCAGCTCCAAACTTTTGAACCTTTTGGTTTCCGTT 185
QY 183 CTCTCTCTCTCTCTCTCACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 242
Db 186 CAGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 208
QY 243 CTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 302
Db 209 -----CCTCCACCTCTCTAAACATGTCCACTGGCTCCCTCAGGAGTGTAGAGACCT 260
QY 303 TCAGAGGTCGAGATGTTGGAATGTGACGGGTTGAAATGGATTTCGAACAAGGAATTTGT 362
Db 261 TCAAGAGTGGAGATGCTGGACTGTGACTCTCCCTGAAAGTGGATCTCAACAGAGTTTG 320
QY 363 GACTTCCACAGCAGACACCGAGGAGAGTCTCAACTCGAGAAATGGGTCTCTCCAGAGGG 422
Db 321 AACTTCCACAGCAGACACCGAGGAGGCTCCAACTCGAGAAAGGCTCTCCAGAGGG 380
QY 423 CCGGGCGGCTGGGCAAGAGGAGGAGCGGCCCAACAGAGAGCCCTTGAGCGGGGT 482
Db 381 TCGCGTGGCTTAGGCAAGAGGAGGAGGCGGCCCACTAAAGAAAGCCGCTCAGCGGGGT 440
QY 483 CAGCCAGAGGGGAAAGCAGGTCCAGGCGCAACGCGCGCAACGCGCAGAGCGGCGCGCAT 542
Db 441 CAGCCAGAGGGGCAAGCAGGTCCAGGCGCAACGCGCGCAATGCTCGTGAAGCGCGCAT 500
QY 543 GCGAGTCTGAGCAAGGCTTCTCTCAGATCAAGACACCTGCGCTGGTGGTGGTGGTGG 602

```

```

Db      501  GCGGCTGCTGAGCAAGGCTTCTCCAGCTCAGACACCCTCGCTGGGTGCCCGCGA 560
Qy      603  CACCAAGCTTCCAGCTGACACGCTCAGCTGAGCTGCGGTCCAGCTACATGCCCACTTGAG 662
Db      561  CACCAAGCTTCCAGCTGACACGCTCAGCTGAGCTGCGGTCCAGCTACATGCCCACTTGAG 620
Qy      663  GCAGATCTGCTGCTAAGCAAAATACGAGACGGGTACATTCACCCGGTCAACCTGACGTG 722
Db      621  GCAGATCTGCTGCTAAGCAAAATACGAGACGGGTACATTCACCCGGTCAACCTGACGTG 680
Qy      723  GCCCTTTATGCTGGCGCGGAAACCCGAGAGTGCCTGAAAGAGTGGTGACCCGAGCGG 782
Db      681  GCCCTTTATGCTGGCGCGGAAACCCGAGAGTGCCTGAAAGAGTGGTGACCCGAGCGG 740
Qy      783  CTTATGTGGAACACCGCTCTGACCTTGGAGGTGGAGTCTGGGAAAGCGCGCTCC 842
Db      741  CTTGTGTGGAACACAGCATCTGACCTTGGAGGTGGAGTCTGGGAAAGCGCTGAGCT-CT 799
Qy      843  GGGGGGAGCGGGCGCGGAAAGCGACCCCTGCGCTCAGTGTCTCTGTCTCTGCTTCCC 902
Db      800  GGGGGGAGTGGCGCGGAAAGCGAGCGACCCCTGCGCTCAGTGTCTCTGTCTCTGCTT 859
Qy      903  CTTGCGAATGCTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 962
Db      860  CTTGCTGCTGCTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 919
Qy      963  TCGTTTCCAAACAGAGGAGATCAATGTACTTACAAAGATCCCATCTTATTACTTTA 1022
Db      920  TCGTTTCCAAACAGAGGAGATCAATGTACTTACAAAGATCCCATCTTATTACTTTA 977
Qy      1023  TTAACCTTCTACCGTGAATGACTCTGCAAGCGCTTGTGCTGCTGCTGCTGCTGCTG 1082
Db      978  TTAACCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1035
Qy      1083  AATATATA-----AATAGATAGACCTTACAA--TGATCTTTTGTACAA 1128
Db      1036  ATATATATATATATATATATATATATATATATATATATATATATATATATATAT 1095
Qy      1129  ATGTTGTAATATAGATCATAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1188
Db      1096  ATGTTGTAATATAGATCATAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1155
Qy      1189  CACTTAAAGATATATATATATATATATATATATATATATATATATATATATAT 1248
Db      1156  CACTCGAAGATATATATATATATATATATATATATATATATATATATATATAT 1215
Qy      1249  TTAATTTGCTAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1272
Db      1216  TTAATTTGCTAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1239

```

```

RESULT 7
BC053525
LOCUS      1246 bp      mRNA      linear      ROD 30-JUN-2004
DEFINITION Mus musculus transcription factor 21, mRNA (CDNA clone MGC:58914
IMAGE:6529241), complete cds.
ACCESSION BC053525
VERSION   BC053525.1 GI:31566109
KEYWORDS  MGC.
SOURCE    Mus musculus (house mouse)
ORGANISM  Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 1246)
Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
Diachenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
Schetz,T.E., Brownstein,M.J., Udwin,T.B., Toshiyuki,S.,
Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,

```

```

TITLE
JOURNAL
PUBMED
REFERENCE
AUTHORS
TITLE
JOURNAL
REMARK
COMMENT
Contact: MGC help desk
Email: cgabbs@mail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: National Institutes of Health Intramural
Sequencing Center (NISC),
Gaithersburg, Maryland;
Web site: http://www.nisc.nih.gov/
Contact: nisc.mc@hghri.nih.gov
Ahter,N., Aylee,K., Beckstrom-Sternberg,S.M., Benjamin,B.,
Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C.P., Brooks,S.,
Dietrich,N.L., Granite,S., Guan,X., Gupta,J., Haghighi,P.,
Hansen,N., Ho,S.-L., Karlins,E., Kwong,P., Lasic,P., Legaspi,R.,
Maduro,Q.L., Masiello,C., Maskeri,B., Mastrian,S.D., McCloskey,J.C.,
McDowell,J., Pearson,R., Stantropop,S., Thomas,P.J., Touchman,J.W.,
Tsurgon,C., Vogt,J.L., Walker,M.A., Wetherby,K.D., Wiggins,L.,
Young,A., Zhang,L.-H. and Green,E.D.
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAC Plate: 108 Row: p Column: 16.
FEATURES
Location/Qualifiers
1. 1246
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="FVB/N"
/clone="MGC:58914 IMAGE:6529241"
/tissue_type="Colon, normal. 5 month old male mouse."
/clone_lib="NCI CGAP_Co24"
/lab_host="DH10B"
/notes="Vector: pCMV-SPORT6"
1. 1246
/gene="Tcf21"
/notes="synonyms: epc, Pod1, capsulin, epicardin, pod-1"
/db_xref="LocusID:21412"
/db_xref="MGI:1202715"
235..772
/gene="Tcf21"
/codon_start=1
/product="transcription factor 21"
/protein_id="AAH53525.1"
/db_xref="GI:31566110"
/db_xref="LocusID:21412"
/db_xref="MGI:1202715"
/translation="MSTGSLSDVEDLQEVMLDCDSLKVDNSKEFGTSNESTEGSNC
KTPSPQGGGLGKRPKPSLGSQGVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQV
KTLPWPVPTDKSLDITLRLASSYIAHLRLQILANDKYENGTHPVNLTWPFVWAGKP
ENDLKVETANRLCGTAS"
ORIGIN

```

```

Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
Fahey,J., Helton,E., Kettman,M., Madan,A., Rodriguez,S.,
Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smalish,D.E.,
Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
12477932
2 (bases 1 to 1246)
Strausberg,R.
Direct Submission
Submitted (09-JUN-2003) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: cgabbs@mail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: National Institutes of Health Intramural
Sequencing Center (NISC),
Gaithersburg, Maryland;
Web site: http://www.nisc.nih.gov/
Contact: nisc.mc@hghri.nih.gov
Ahter,N., Aylee,K., Beckstrom-Sternberg,S.M., Benjamin,B.,
Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C.P., Brooks,S.,
Dietrich,N.L., Granite,S., Guan,X., Gupta,J., Haghighi,P.,
Hansen,N., Ho,S.-L., Karlins,E., Kwong,P., Lasic,P., Legaspi,R.,
Maduro,Q.L., Masiello,C., Maskeri,B., Mastrian,S.D., McCloskey,J.C.,
McDowell,J., Pearson,R., Stantropop,S., Thomas,P.J., Touchman,J.W.,
Tsurgon,C., Vogt,J.L., Walker,M.A., Wetherby,K.D., Wiggins,L.,
Young,A., Zhang,L.-H. and Green,E.D.
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAC Plate: 108 Row: p Column: 16.
FEATURES
Location/Qualifiers
1. 1246
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="FVB/N"
/clone="MGC:58914 IMAGE:6529241"
/tissue_type="Colon, normal. 5 month old male mouse."
/clone_lib="NCI CGAP_Co24"
/lab_host="DH10B"
/notes="Vector: pCMV-SPORT6"
1. 1246
/gene="Tcf21"
/notes="synonyms: epc, Pod1, capsulin, epicardin, pod-1"
/db_xref="LocusID:21412"
/db_xref="MGI:1202715"
235..772
/gene="Tcf21"
/codon_start=1
/product="transcription factor 21"
/protein_id="AAH53525.1"
/db_xref="GI:31566110"
/db_xref="LocusID:21412"
/db_xref="MGI:1202715"
/translation="MSTGSLSDVEDLQEVMLDCDSLKVDNSKEFGTSNESTEGSNC
KTPSPQGGGLGKRPKPSLGSQGVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQV
KTLPWPVPTDKSLDITLRLASSYIAHLRLQILANDKYENGTHPVNLTWPFVWAGKP
ENDLKVETANRLCGTAS"
ORIGIN

```

		Query Match		60.9%; Score 774.4; DB 10; Length 1246;	
		Best Local Similarity		79.4%; Pred. No. 1.8e-147;	
		Matches 1019; Conservative		0; Mismatches 201; Indels 64; Gaps 6;	
Qy	3	TACGGCCAGACTCTGGAGTGGGAAACAGAGAGCCGGTTCCTCTGCTGCAAGAGTCT	62		
Db	13	TACGGCCAGACTCTGGAGTGGGAAACAGAGTGGTTCCTCTGCTGCAAGAGTCT	72		
Qy	63	CGGGGTTCTCTTCAACACTCTCGAAGCGGAAAGGTTGTGAGACCCCAACAGACCCCA	122		
Db	73	TGTGATCGTTTTCACGACTCTGTAAAGGAAAGCTGGGGGACGGACTGCCAGATCCCA	132		
Qy	123	ACTCCAGCTCCAGCAGAGAGTGGCTGGCCACACTCGGAGGCTCTTGGTTTCAGGGT	182		
Db	133	CCTCAACCCCAACACAGAGTGGCAGCTCCAAACTTTGAGACCTTTTCGTTTCGGTTT	192		
Qy	183	CTCTCTGTCTCTCTCAGCCCTCTTCCCTCGCTTCTCTGTCTCTCTCTCTCTCTCT	242		
Db	193	CAGCGTCT	215		
Qy	243	CT	302		
Db	216	-----CCTCCACCTCTCTAAACATGTCACCTGGCTCTCTCTCTCTCTCTCTCT	267		
Qy	303	TCAAGAGTGGAGATGTTGAATGTGACGGGTGAAATGGATGGAACAAAGGAATTTGT	362		
Db	268	TCAAGAGTGGAGATGTTGAATGTGACGGGTGAAATGGATGGAACAAAGGAATTTGT	327		
Qy	363	GACTTCAACAGAGACAGCAGAGAGAGTCCAACTGCGAGATGGTGTCTCCCAAGAGG	422		
Db	328	AACCTTCAACAGAGACAGCAGAGAGAGTCCAACTGCGAGATGGTGTCTCCCAAGAGG	387		
Qy	423	CCCGCGGCTGGCAAGAGAGAGAGGCGCCACCAAGAGAGAGCCCTCTCAGCGGGT	482		
Db	388	TCGCGTGGCTAGCAAGAGAGAGAGGCGCCACCAAGAGAGAGCCCTCTCAGCGGGT	447		
Qy	483	CAGCCAGGAGGGAAGCAGTCCAGCGCAACCGCGCAACCGCGAGAGCGGGCCGGAT	542		
Db	448	CAGCCAGGAGGGAAGCAGTCCAGCGCAACCGCGCAACCGCGAGAGCGGGCCGGAT	507		
Qy	543	GCAGTGTGAGCAAGGCTTCTCAGACTCAAGACACCCCTGCTGGTGGTGGTGGTGGT	602		
Db	508	GCAGTGTGAGCAAGGCTTCTCAGACTCAAGACACCCCTGCTGGTGGTGGTGGTGGT	567		
Qy	603	CACCAAGCTCTCAAGCTGACACGCTCAGGCTGGCTCCAGCTACATCGCCACCTTGAG	662		
Db	568	CACCAAGCTCTCAAGCTGACACGCTCAGGCTGGCTCCAGCTACATCGCTCACTTAAG	627		
Qy	663	GCAGTGTGAGCAAGGCTTCTCAGACTCAAGACACCCCTGCTGGTGGTGGTGGTGGT	722		
Db	628	GCAGTGTGAGCAAGGCTTCTCAGACTCAAGACACCCCTGCTGGTGGTGGTGGTGGT	687		
Qy	723	GCCTTTATGTTGGCGGAAACCCGAGAGTGCCTGAAAGAGTGGTGCACCGCGAGCG	782		
Db	688	GCCTTTATGTTGGCGGAAACCCGAGAGTGCCTGAAAGAGTGGTGCACCGCGAGCG	747		
Qy	783	CTTATGTTGGAACCAACCGCTCTCTGTTCCACCGGAGAACCTTTACACGAGAGAGAT	842		
Db	748	CTTATGTTGGAACCAACCGCTCTCTGTTCCACCGGAGAACCTTTACACGAGAGAGAT	806		
Qy	843	GGGGGAGCGGGCCCGGGAAGCGGACCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCT	902		
Db	807	GGGGGAGCGGGCCCGGGAAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	866		
Qy	903	CCTCGCAATCT	962		
Db	867	CCTTGCATCT	926		
Qy	963	TCGTTTCCAAACAGAGAGATCAATTTGTTTACAAAGTTTCCATCTATTACTTTA	1022		
Db	927	TCGTTTCCAAACAGAGAGATCAATTTGTTTACAAAGTTTCCATCTATTACTTTA	984		

Qy	1023	TTAACTTCTACCGTGAATGACTCTGCAAGCCTTGTGGTCCAGTGC	1082		
Db	985	TTAACTTCTGCCAATGAATGA--CTGCAAGCCTGGCTGGTCCAGTGC	1042		
Qy	1083	AAATATATA-----AATAGATAAGAGCCCTATCAA--TGTATCTTTTGTACAAT	1128		
Db	1043	ATATATATATATATATAAATAGATAAGAGAGCCCTATCAATGTGTATCTTTTGTACAAT	1102		
Qy	1129	ATGTTGTAATAATCTAGATCATAGGATAGCTGACTTTTGACAGTGCACATTTATAAAGTAATT	1188		
Db	1103	ATGTTGTAATAATCTAGATCATAGGATAGCTGACTTTTGACAGTGCACATTTATAAATAAAGT	1162		
Qy	1189	CACCTTAAAGATATATATTTTTTTTCAACCAAGTTTTGCTACTTTTTGAAAAATAAAATCTTTCT	1248		
Db	1163	CACCTGGAAGAAATATATTTTTTTTCAACCAAGTTTTGCTACTTTCCGAAAATAAATCTTTCT	1222		
Qy	1249	TTATATTGCTTAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA	1272		
Db	1223	TTATATTGAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA	1246		
RESULT 8					
AF036945		1267 bp		mRNA	linear
LOCUS		Mus musculus capsulin mRNA, complete cds.			
DEFINITION		AF036945			
ACCESSION		AF036945			
VERSION		AF036945.1 GI:2695694			
KEYWORDS		Mus musculus (house mouse)			
SOURCE		Mus musculus			
ORGANISM		Mus musculus			
REFERENCE		1 (bases 1 to 1267)			
AUTHORS		Lu, J., Richardson, J.A. and Olson, E.N.			
TITLE		Capsulin: a novel bHLH transcription factor expressed in epicardial progenitors and mesenchyme of visceral organs			
JOURNAL		Mech. Dev. 73 (1), 23-32 (1998)			
MEDLINE		98213628			
PUBMED		9545521			
REFERENCE		2 (bases 1 to 1267)			
AUTHORS		Lu, J. and Olson, E.N.			
TITLE		Direct Submission			
JOURNAL		Submitted (04-DEC-1997) Molecular Biology and Oncology, UT Southwestern Medical Center, 6000 Harry Hines Blvd., N8.510, Dallas, TX 75235-9148, USA			
FEATURES		Location/Qualifiers			
source		1..1267			
		/organism="Mus musculus"			
		/mol_type="mRNA"			
		/db_xref="taxon:10090"			
		264..803			
		/function="bHLH transcription factor"			
		/codon_start=1			
		/product="capsulin"			
		/protein_id="AAC23537.1"			
		/db_xref="GI:2695695"			
		/translation="MSTGSLSDVDELDQEVEMLDSDSLKVDNKEPFGTSNESTEGSNC			
		ENSPQKRGRLGKRRKAPTKPSPLSGVSGEKQVORNAARERARMLVSKAFSL			
		KTLTPVPFPDPTKLSDKLDLRLASSYTAHLRLAILANDKYENGYIHPVNLTPFWVAKP			
		ENDLKEVVTANRLCGTTAS"			
CDS					
ORIGIN					
Query Match 60.1%; Score 764.4; DB 10; Length 1267;					
Best Local Similarity 79.2%; Pred. No. 1.9e-145;					
Matches 1009; Conservative 0; Mismatches 201; Indels 64; Gaps 6;					
Qy	3	TACGGCCAGACTCTGGAGTGGGAAACAGAGAGCCGGTTCCTCTGCTGCAAGAGTCT	62		
Db	44	TACGGCCAGACTCTGGAGTGGGAAACAGAGAGCCGGTTCCTCTGCTGCAAGAGTCT	103		
Qy	63	CGGGTTCCTTTCACAACTCTTCGGAAGGGGAAAGGTTGTGAGACCCCAACAGACCCCA	122		
Db	104	TGTGATCGTTTTCACGACTCTGTAAAGGAAAGCTGGGGGACGGCACTGCCAGATCCCA	163		

```
123 ACTCCAGCTCCACGAGAGTGGCTGCGCCACACTCGGAGGCGCTCTTGTTTCAGGGT 182
164 CCTCAACCCCAACACAGAGAGTGGCAGCTCCAACTTTTGAGACCCCTTCGTTTCGGTTT 223
183 CTCCTCTCTCTCTCTCTCACCCTCTTCCTCGCTTTCTCTCTCTCTCTCTCTCTCT 242
224 CAGCGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 246
243 CTCCTCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 302
247 -----CTCTCACCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 298
303 TCAAGAGGTGGAGATGTTGGAATGTGACGGGTTGAAAATGGATTCGAAACAAGGAATTTGT 362
299 TCAAGAGGTGGAGATGCTGACTGTGACTCCCTGAAAGTGACTCCCAACAAGAGATTGG 358
363 GACTTCAACGAGAGACCCGAGGAGAGCTCCAATCGGAGAAATGGGTCTCTCCCAAGGG 422
359 AACTTCAACGAGAGACCCGAGGAGGCTCCAACTCGAGAACGGGTCTCCACAGAAGGG 418
423 CCGCGCGGCTGGGCAAGAGGAGGAGGCGGCCACCAAGAAGAGCCCTCTGAGCGGGT 482
419 TCGCGTGGCTTAGGCAAGAGGAGGAGGCGGCCACTTAGAAAAGCCGCTCAGCGGGT 478
483 CAGCAGAGGGGAGAGAGTCCAGGCGCAACCGCGCAACCGCGGAGAGCGGCCCGCAT 542
479 CAGCAGAGGGGCAAGCAGGTCCAGGCGCAACCGCGGCAATGCTCTGAGCGGCGCCGCAT 538
543 GCGAGTCTCAGCAAGGCTCTCCAGACTCAAGACCCCTCCCTGGGTGCCCCCGCA 602
539 GCGGTGCTGAGCAAGGCTCTCCAGGCTCAAGACCCCTCCCTGGGTGCCCCCGCA 598
603 CACCAAGCTCTCAAGCTGACACGCTCAGGCTGGGTCCAGTACATGCCACATGAG 662
599 CACCAAGCTCTCAAGCTGACACGCTCAGGCTGGGTCCAGTACATGCCACATGAG 658
663 GCAGATCTGGCTTAACGACAAATACGAGACGGGTACATTCACCGGTCAACCTGACGTG 722
659 GCAGATCTGGCTTAACGACAAATACGAGACGGGTACATTCACCGGTCAACCTGACGTG 718
723 GCCTTTATGTTGGTGGCGGGAACCCGAGAGTACCTGAAAGAAAGTGGTGACCGCGAGCG 782
719 GCCTTTATGTTGGTGGCGGGAACCCGAGAGTACCTGAAAGAAAGTGGTGACCGCGAGCG 778
783 CTTATGTGAAACACCGCGTCTCTGACTTGGAGGTGCGAGTCTGGGAAAGGCGGTCC 842
779 CTTGTGTGAACTACAGCATCTCTGACCTTGGAGGTGCGAGTCTGGGAAACCGTGAGCT-CT 837
843 GGGGGAGCGGCGCGGGAAGCGGACCCCTGCGCTCAGTCTCTCTCTCTCTCTCTCTCT 902
838 GGGGGAGTGGGCGGGAAGGTGACGACCCCTGCGCTCTCTCTCTCTCTCTCTCTCTCT 897
903 CTTGCAATGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 962
898 CTTTGCATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 957
963 TCGTTTCCAAACGAGGAGATCAATGTACTTACAAAGATCCCACTCTATTATTCTTTA 1022
958 TCGTTTCCAAACGAGGAGACCACTGTACTTAC--CGATTTCTATTTATTATTCTTTA 1015
1023 TTAACCTTCTACCGTGAATGACTCTGCAAGCCTTGTGCTGCTCCAGTCAATATGTAATTAT 1082
1016 TTAACCTTCTGCAATGATGA--CTGCAAGCTCTGCTGCTGCTCCAGTCAATATGTAATTAT 1073
1083 AATATATA-----AATAGATAGAGCCCTATCA--TGATCTTTTGTACAT 1128
1074 ATATATATATATATATAAATAGATAGAGCCCTATCAATGTGTATCTTTTGTACAT 1133
1129 ATGTGTAAATGTAGATCATAGATAGTGTGCTTTGACAGTCAATTTTATAAGTAATT 1188
1134 ATGTTGTAAATGTAGATCATAGATAGTGTGCTTTGACAGTCAATTTTATAAATAGT 1193
```

---

```
QY 1189 CACTTAAAGATATATATTTTTTCAACAAGTTTGTCTACTTTTGAAATAAATCTTCT 1248
Db 1194 CACTCGAAGAAATATATTTTTTCAACAGGTTTGTCTACTTCCGAAATAAATCTTCT 1253
QY 1249 TTATATTGCTAAAA 1262
Db 1254 TTATATTGAAAAAA 1267

RESULT 9
AB009453
LOCUS Mus musculus mRNA for nephonadin, complete cds. linear ROD 11-DEC-1997
DEFINITION AB009453
ACCESSION AB009453
VERSION AB009453.1 GI:2696117
KEYWORDS nephonadin.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 1220)
AUTHORS Tamura,M. and Nakatsuji,N.
TITLE Identification of Nephonadin, a novel basic helix-loop-helix gene
JOURNAL Published Only in Database (1997)
REFERENCE 2 (bases 1 to 1220)
AUTHORS Tamura,M. and Nakatsuji,N.
TITLE Direct Submission
JOURNAL Genet. Mammalian Development; Yata 1111, Mishima, Shizuoka 411, Japan (E-mail:matamura@lab.nig.ac.jp, Tel:+81-559-81-6832, Fax:+81-559-81-6828)
FEATURES
source
1..1220
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
223..762
/function="transcription factor"
/note="basic helix-loop-helix gene"
CDS
223..762
/codon_start=1
/product="nephonadin"
/protein_id="BAA23883.1"
/db_xref="GI:2696118"
/translation="MSTGSLSDVEDLQEVMLDCDSLKVDNKEFTSNSTERSGNC
ENGSPQKRGGLGKRRKAPTKSPSGQKQVQRNAAARARMRVLRKAPSRLL
KTLFPPFPPTKLKLDLRLASSYIAHLRQILANDKYENGVIHPVNLTPFVWAGK
ENDLKEVVVTANRLCGTTAS"
polyA_site
1220
/note="17 A nucleotides"
ORIGIN
Query Match 59.9%; Score 761.6; DB 10; Length 1220;
Best Local Similarity 79.3%; Pred. No. 7.1e-145;
Matches 1005; Conservative 0; Mismatches 199; Indels 64; Gaps 6;
QY 3 TACGCCACGACTCTGGAGTGGGAAACAGAGAGCGGTTCTCTCTGTCGAGAGTCT 62
Db 3 TAAGCCAGACTCCAGTGTGGGAGATAGAGACTGGTTTCTCTGCGGGATCTCT 62
QY 63 CGGGTTCTTCTTCTCAAACTCTCGAAGGGGAAAGGTTGTGAGACCCCAACAGACCCCA 122
Db 63 TGTGATCTGTTTCCACGACTCTGTAAGGGGAAAGCTGGGGACGGGACTGCCAGATCCA 122
QY 123 ACTCAGTCTCCACGAGAGTGGTGGCCACACTCGGAGGCTCTTGGTTTCAGGTT 182
Db 123 CTTCAACCCCAACAGAGTGGCAGCTCCAAACTTTTGAGACCTTTTCCGTTT 182
QY 183 CTCCTGTCTCTCTCTCACCCTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 242
Db 183 CAGCGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 205
QY 243 CTCTCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 302
```

206 -----CCTCACCTCTCTAAACATGTCCACTGCTCCCTCAGCGATGTAGAAGACCT 257  
 QY TCAGAGGTGGAGATGTTGGAATGTGACGGTTGAAATGGATTGGAACAAGGAATTTGT 362  
 Db TCAGAGGTGGAGATGTTGGAATGTGACGGTTGAAATGGATTGGAACAAGGAATTTGT 317  
 QY GACTTCCAGAGAGACACGAGAGAGCTTCAAACGCGAGATGGGTCTCCCAAGAGG 422  
 Db AACTTCCAGAGAGACACGAGAGAGCTTCAAACGCGAGATGGGTCTCCCAAGAGG 377  
 QY CCGCGGGGCTGGCAGAGAGAGAGAGCGGCCACACAGAGAGAGCGCCCTCAGCGGGT 482  
 Db TCAGAGGTGGAGATGTTGGAATGTGACGGTTGAAATGGATTGGAACAAGGAATTTGT 437  
 QY CAGCCAG 542  
 Db CAGCCAG 497  
 QY GCGAGTGTGAGCAAGGCTTCTCAGACTCAAGACACCGCTGCGCTGGTGGTGGTGGTGG 602  
 Db GCGGTGTGAGCAAGGCTTCTCAGACTCAAGACACCGCTGCGCTGGTGGTGGTGGTGG 557  
 QY CACCAAGCTCTCAAGCTGACACGCTCAGGCTGGGCTCAGCTACATCGCCCACTTGAG 662  
 Db CACCAAGCTCTCAAGCTGACACGCTCAGGCTGGGCTCAGCTACATCGCTCACTTAAG 617  
 QY GCGAGTGTGAGCAAGGCTTCTCAGACTCAAGACACCGCTGCGCTGGTGGTGGTGGTGG 722  
 Db GCAGATCTCTGCCCAACACAAAGTACGAGAACCGGTACATTCACCGGTCAACTGACGTG 677  
 QY GCGCTTTATGTCGCGGGAACCCGAGAGTACCTGAAAGAGTGTGACCGCGAGCGG 782  
 Db GCGCTTTATGTCGCGGGAACCCGAGAGTACCTGAAAGAGTGTGACCGCGAGCGG 737  
 QY CTTATGTGGAACACACCGCTCTGACCTTGGAGTGTGAGTGTGGAAGAGGCGCGCTCC 842  
 Db CTTGTGTGGAACACACCGCTCTGACCTTGGAGTGTGAGTGTGGAAGAGGCGCGCTCC 796  
 QY GGGGGAGCGGGCGCGGAGCGGACCGCTGCGCTCAGTGTCTCTGCTCTGCTTCC 902  
 Db GGGGGAGCGGGCGCGGAGCGGACCGCTGCGCTCAGTGTCTCTGCTCTGCTTCC 856  
 QY CTTGCAATGCTCTCTCTGCTCCACCGCGAGAACACTTTTCAACAGAGAGAGAGAT 962  
 Db CTTGCAATGCTCTCTCTGCTCCACCGCGAGAACACTTTTCAACAGAGAGAGAGAT 916  
 QY TCGTTTCCAAACAGAGAGAGATCAATGTGACTTCAAGATTTCCCATCTATTACTTTA 1022  
 Db TCGTTTCCAAACAGAGAGAGATCAATGTGACTTCAAGATTTCCCATCTATTACTTTA 974  
 QY TTAACCTTCCGCAATGATGA--CTGCAAGCTTGGTGGTCCAGTGCATATGTAATTA 1082  
 Db TTAACCTTCCGCAATGATGA--CTGCAAGCTTGGTGGTCCAGTGCATATGTAATTA 1032  
 QY AATATATA-----AATAGATAGAGCCTATCAA--TGATCTTTTGTACAA 1128  
 Db AATATATA-----AATAGATAGAGCCTATCAA--TGATCTTTTGTACAA 1092  
 QY ATGTTGTAATGTAGATCATAGATAGCTGCTTTGACAGTGCACATTTTATAAGTAAT 1188  
 Db ATGTTGTAATGTAGATCATAGATAGCTGCTTTGACAGTGCACATTTTATAAGTAAT 1152  
 QY CACTTAAAGATATATATTTTTTCAACAGTTTGTCTACTTTTGAATAAATCTTTCT 1248  
 Db CACTTAAAGATATATATTTTTTCAACAGTTTGTCTACTTTTGAATAAATCTTTCT 1212  
 QY TTAATTTG 1256  
 Db TTAATTTG 1220

LOCUS AF029753 1217 bp mRNA linear ROD 25-FEB-2000  
 DEFINITION Mus musculus basic helix-loop-helix factor Cor1 mRNA, complete cds.  
 ACCESSION AF029753  
 VERSION AF029753.1 GI:2613018  
 KEYWORDS  
 SOURCE Mus musculus (house mouse)  
 ORGANISM  
 REFERENCE  
 1 (bases 1 to 1217)  
 AUTHORS Hidai,H., Bardales,R., Goodwin,R., Quertermous,T. and Quertermous,E.E.  
 TITLE Cloning of capsulin, a basic helix-loop-helix factor expressed in progenitor cells of the pericardium and the coronary arteries Mech. Dev. 73 (1), 33-43 (1998)  
 JOURNAL  
 MEDLINE  
 PUBMED  
 98213629  
 9545526  
 REFERENCE  
 2 (bases 1 to 1217)  
 AUTHORS Quertermous,T., Quertermous,E.E., Hidai,H., Fadel,B., Boutet,S.C. and Bardales,R.  
 TITLE Direct Submission  
 JOURNAL Submitted (10-OCT-1997) Department of Medicine, Division of Cardiology, Stanford University, Falk Bldg, 300 Pasteur Drive, Stanford, CA 94305, USA  
 FEATURES  
 Location/Qualifiers  
 1..1217  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:10090"  
 /tissue\_type="pericardium and coronary arteries"  
 1..1217  
 /gene="Cor1"  
 213..752  
 /gene="Cor1"  
 /function="transcription factor"  
 /codon\_start=1  
 /product="basic helix-loop-helix factor Cor1"  
 /protein\_id="AAB84256.1"  
 /db\_xref="GI:2613018"  
 /translation="MSTGSLSDVEDLQEVMLDCSLKVDNKEFGTSNESTEGSNC  
 ENSFQKRGKGLKRRKAPKPKPLSGVSGKQVQRNANARPERARMLVSKAFRL  
 KTLPLVPDPDKLSKLDLRLASSYIAHLNLIQLANDKYNENYHPLNLTWPNVAKGP  
 ENDLKEYVTANRLCGTTAS"  
 ORIGIN  
 Query Match 59.7%; Score 759.6; DB 10; Length 1217;  
 Best Local Similarity 79.2%; Pred. No. 1.8e-144;  
 Matches 1003; Conservative 0; Mismatches 199; Indels 64; Gaps 6;  
 QY 12 GACTCTGGAGTGGGAAACAGAGAGCGGTTCTCTGCTGCGAGAGTCTCTGGGTTCC 71  
 Db 2 GACTCCAGTGTGGGAGATAGACACTGGTTTCCCTGTGCGGGATCCCTTGTGATCGT 61  
 QY 72 TTCTCAACATCTCGAAGGGGAAGGTTGTGAGACCCCAACAGACCCCACTCCAGCT 131  
 Db 62 TTTCCAGACTCTCTAAGGGAAGAGCTGGGGAGCGGACTGCCAGATCCCACTCAACC 121  
 QY 132 CCCAGCAGAGGTGGTGGCGCCACACTCGGAGAGCCCTCTTGGTTTCAGGCTCTCTGTC 191  
 Db 122 CAACACAGAGTGGCAGCTCCAAACTTTTGAGACCCCTTTCGTTTCGTTTCAGCGTCTC 181  
 QY 192 TCTCTCTACCCCTC 251  
 Db 182 TC 196  
 QY 252 GTCCACTCCCCCAACATGTCCACCGGCTCCCTCAGCGATGTGGAGACCTTCAAGAGGT 311  
 Db 197 CTCACCTCTCTAAGCATGTCCACTGGCTCCCTCAGCGATGTAGAGACCTTCAAGAGGT 256  
 QY 312 GGAGATGTGGAATGTGACGGGTTGAAAATCGAATTCGAACAGGAATTTGTGACTTCCAA 371  
 Db 257 GGAGATGTGGAATGTGACGGGTTGAAAATCGAATTCGAACAGGAATTTGTGACTTCCAA 316

QY 372 CGAGAGCACCGAGAGAGTCCAACTGCGAGAAATGGTCTCCCAAGAGGCGCGGG 431  
 Db 317 CGAGAGCACCGAGAGAGGCTCCAACTGCGAGAAATGGTCTCCCAAGAGGCGCGGG 376  
 QY 432 CTTGGGCAAGAGAGAGGCGCCCAACCAAGAGAGGCGCGGCTGAGCGGGTACGACAGA 491  
 Db 377 CTTAGGCAAGAGAGAGGCGCCCAACCAAGAGAGGCGCGGCTGAGCGGGTACGACAGA 436  
 QY 492 GGGGAAGCAGGTCAGGCGCAACCGCCCAACCGCGAGAGCGGCGCGGATCGAGTGT 551  
 Db 437 GGGCAAGCAGGTCAGGCGCAACCGCCCAACCGCGAGAGCGGCGCGGATCGAGTGT 496  
 QY 552 GAGCAGGCTTCTCAGAGCTCAGAGCAACCGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCT 611  
 Db 497 GAGCAGGCTTCTCAGAGCTCAGAGCAACCGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCT 556  
 QY 612 CTCAGAGCTGAGACAGCTCAGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCT 671  
 Db 557 CTCAGAGCTGAGACAGCTCAGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCT 616  
 QY 672 GGCTAAGCAAAATAGAGAGCGGTAATTCACCGGCTCAACCTGAGCTGAGCTGCGGCTTAT 731  
 Db 617 GGGCAAGCAAAATAGAGAGCGGTAATTCACCGGCTCAACCTGAGCTGAGCTGCGGCTTAT 676  
 QY 732 GGTGGCGGGAACCGAGAGTCACTGAAAGAGTGTGACCGGAGCGGCTTATGTCG 791  
 Db 677 GGTGGCGGGAACCGAGAGTCACTGAAAGAGTGTGACCGGAGCGGCTTATGTCG 736  
 QY 792 AACACCGGCTCTGACCTTGGAGTGGAGTCTGGGAAAGCGCGCTCCCGGGGGAGC 851  
 Db 737 AACTAGCATCTGACCTTGGAGTGGAGTCTGGGAAAGCGCGCTCCCGGGGGAGC 795  
 QY 852 GGGCGCGGGAAGGCGGAGCGGCTGCGGCTGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 911  
 Db 796 GGGCGGGAAGGCTGAGCGGCTGCGGCTGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 855  
 QY 912 GCTCCT 971  
 Db 856 GCTCCT 915  
 QY 972 AACGAGAGGATCAATGTACTTACAAAGATCCCACTATTAATTAATTAATTAATTAATTAAT 1031  
 Db 916 AACCAAGGAGACCACTGTACTTAC--CGAATCTCATCTATTAATTAATTAATTAATTAAT 973  
 QY 1032 ACCGTGAATCACTCTGCAAGCCTTGTGCTGCAAGTCAATATGTAATTAATTAATTAATTAAT 1091  
 Db 974 GCCATGATGA--CTGCAAGCCTGCTGCTGCAAGTCAATATGTAATTAATTAATTAATTAAT 1031  
 QY 1092 -----AATGATAAGAGCCTATCAA--TGATCTTTTGTACAAATATGTTGTAA 1137  
 Db 1032 TTATATAAATAGATAAGAGAGCCTATCAATGTGTATCTTTTGTACAAATATGTTGTAA 1091  
 QY 1138 AATGTAGATCATAGGATAGCTGCTTGTGACGTCACATTTATAAGTAATTAATTAATTAATTAAG 1197  
 Db 1092 AATGTAGATCATAGGATAGCTGCTTGTGACGTCACATTTATAAGTAATTAATTAATTAATTAAG 1151  
 QY 1198 ATATATATTTTTTCAACCAAGTTTTGCTACTTTTCAAAATTAATCTTTCTTTATATTCG 1257  
 Db 1152 AAATATATTTTTTCAACCAAGTTTTGCTACTTTCCGAAATTAATCTTTCTTTATATTCG 1211  
 QY 1258 TAAAAA 1263  
 Db 1212 AAAAAA 1217

RESULT 11  
 AF047418  
 LOCUS AF047418 1202 bp mRNA linear ROD 06-OCT-1998  
 DEFINITION Mus musculus epicardin mRNA, complete cds.  
 ACCESSION AF047418  
 VERSION AF047418.1 GI:2921850  
 KEYWORDS Mus musculus (house mouse)

ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE 1 (bases 1 to 1202)  
 AUTHORS Robb, L., Mifusud, L., Hartley, L., Biben, C., Copeland, N.G.,  
 Gilbert, D.J., Jenkins, N.A. and Harvey, R.P.  
 TITLE epicardin: A novel basic helix-loop-helix transcription factor gene  
 expressed in epicardium, branchial arch myoblasts, and mesenchyme  
 of developing lung, gut, kidney, and gonads  
 Dev. Dyn. 213 (1), 105-113 (1998)  
 JOURNAL 98401941  
 MEDLINE 9733105  
 PUBMED 9733105  
 REFERENCE 2 (bases 1 to 1202)  
 AUTHORS Robb, L. and Mifusud, L.  
 TITLE Direct Submission  
 JOURNAL Submitted (10-FEB-1998) Haematology and Cancer, Walter and Eliza  
 Hall Institute of Medical Research, PO Royal Melbourne Hospital,  
 Melbourne, VIC 3050, Australia  
 FEATURES  
 Location/Qualifiers  
 1..1202  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="Balb/c"  
 /db\_xref="taxon:10090"  
 199..738  
 /function="basic helix-loop-helix transcription factor"  
 /note="expressed in embryonic epicardium and  
 mesenchyme-derived tissues of lung, gut, kidney, and  
 gonad"  
 /codon\_start=1  
 /product="epicardin"  
 /protein\_id="AAC62531.1"  
 /db\_xref="GI:2921851"  
 /translation="MSTGSLSDVEDLQEVMLDCDSLKVDNSKEFGTSNESTEGSNC  
 ENGSPQGRGLGRKAPTKPGLSGVSGQKQVQRNANARARMLVLSKAFSL  
 KTLPLWPDPDTKLKSLDLRLASSYTAHLRQILANDKYENGYIHPVNLTPFWVAGRP  
 ENDLKVVVTANRLCGTTAS"  
 ORIGIN  
 Query Match 59.13; Score 752; DB 10; Length 1202;  
 Best Local Similarity 79.13; Pred. No. 6.4e-143;  
 Matches 993; Conservative 0; Mismatches 195; Indels 64; Gaps 6;  
 QY 25 GGGAAACAGAGAGCGGTTCTCTGCTGCAGAGTCTCTCGGGGTTCTCTCACAATCT 84  
 Db 1 GGGAGATAGACATGTTTCCCTGTGCGGGATCCCTTGATCGTTTCCACGACTCT 60  
 QY 85 GCGAAGGGGAAGGGTGTGAGACCCAAACAGACCCCAACTCCAGTCCCGAGGAGGT 144  
 Db 61 GTAAAGGGAAGCTGGGGACGGGACTGCCAGATCCCACTCAAAACCAACACAGAGT 120  
 QY 145 GGCTGCGCCACATCGGGAGGCTCTTGGTTTCAAGGCTCTCTGCTCTCTCTCACCCT 204  
 Db 121 GGCAGCTCCAACTTTTGAGACCCCTTTCGTTTCGTTTACGGTCTCTCTCTCTCTCT 180  
 QY 205 CTTCTCTGCTTCT 264  
 Db 181 C-----CCTCCACCTCTA 195  
 QY 265 AACATGTCACCGGCTCCCTCAGGATGTGGAGACCTTCAAGAGGTGGAGATTTGGAA 324  
 Db 196 AACATGTCACCTGCTCCCTCAGGATGTGGAGACCTTCAAGAGGTGGAGATTTCTGGAC 255  
 QY 325 TGTGACGGGTTGAAAATGGATTCCGAACAAGGAATTTGTGACTTCCAAACGAGACACCGAG 384  
 Db 256 TGTGACTCCCTGAAAGTGGACTCCAAACAGGAGTTTGGAACTTCCACGAGAGACCGAG 315  
 QY 385 GAGAGCTTCAACTCGGAGATGGGTCTCCCAAGAGGCGCGGCGCTTGGGCAAGAG 444  
 Db 316 GAGGCTTCAACTCGGAGAGCGGCTCTCCACAGAGGCTCGCGTGGCTTGGCAAGAG 375  
 QY 445 AGGAAGGCGGCCCAACAGAGAGAGCCCTGAGCGGGGTGAGCCAGGAGGGGAGACAGTGC 504





```
/product="bA373A10.1 (transcription factor 21)"
/notes="match: cDNAs: Em:AB009453 Em:AF036945 Em:AF029753
Em:AF035717 Em:AF047418 Em:AF035718 Em:AF047419
Em:AF087036 Em:AF060154 Em:AF108216 Em:M99367 Em:AF061752
Em:Y11255 Em:L35922 Em:AF109012 Em:M86923 Em:Y09596
Em:AB007959 Em:U18658
match: ESTs: Em:AW313610 Em:W73859 Em:AA880760 Em:A1927658
Em:R22321 Em:AW357120 Em:AW432075 Em:AW190108 Em:AW025547
Em:AW242648 Em:AW548204 Em:A1971939 Em:W73825 Em:R65626
Em:W08124 Em:AW558915 Em:AW086478 Em:AA154683 Em:AA028184
Em:AA027032 Em:AA335365 Em:A1478874 Em:A1578558 Em:N95204
Em:R63028 Em:AV374791 Em:W21552 Em:AA028092 Em:N66932
Em:AA335824 Em:A1712521 Em:AV114060 Em:C21328 Em:A185851
Em:H03852 Em:AA834021 Em:A1721064 Em:AV365142 Em:AV366354
Em:AA027107 Em:AV311062 Em:AV365865 Em:AA341136
Em:A1336106 Em:AA905950 Em:AV351072 Em:A1178188
Em:AA219936 Em:AW197130 Em:HS1772 Em:AA340993 Em:AV351557
Em:D45696 Em:H95174"
/evidence=not experimental
repeat_region
13096..13165
/notes="135 copies 2 mer to 85% conserved"
join(13182..13631,15497..15586)
/genes="TCF21"
/notes="match: proteins: Tr:Q62523 Tr:Q88940 Tr:P79782
Tr:Q75946 Tr:Q43545 Tr:Q64124 Tr:Q73615 Sw:Q13562
Tr:Q91418 Tr:Q61100 Tr:Q942E5 Tr:Q9YH81 Tr:Q9YH82
Tr:Q90690 Sw:Q79920 Sw:Q60867 Sw:P24899 Tr:Q9W6C6
Tr:Q9PT50 Sw:Q60430 Tr:Q9PTE4 Sw:Q64289 Sw:P79765
Tr:Q60682 Tr:Q70483 Tr:Q9YHA3 Tr:Q95300 Tr:Q61246
Tr:Q98939 Tr:Q9YH69 Tr:P797831 Sw:Q91616 Sw:P79766
Tr:P797833"
/codon_start=1
/evidence=not experimental
/product="bA373A10.1 (transcription factor 21)"
/protein_id="CAC26822.1"
/db_xref="GI:12339805"
/db_xref="GOA:O43680"
/db_xref="Swiss-Prot:O43680"
/translations="MSGISDVIEDLQVEVMECDGLKSDNSKEFYTSNTEESSNC
ENGSPKGRGLGKRKAPTKSGPLSGVQEGQVQVNAANARERARMLVLSKFSRL
KTLPLWPPDPTKLSKLDLRLASSYIAHLRLQILANDKYENGVIHPVNLTPFPWVAGK
ESDLKEVVTASRLCGTAS"
13286..13864
/genes="TCF21"
/notes="CpG island"
/evidence=not experimental
13326..13636
/genes="TCF21"
/notes="match: GSS: Em:AL212694 Em:AL254874"
complement(13513..13636)
/notes="match: GSS: Em:AL214658"
15337..15744
/genes="TCF21"
/notes="CpG island"
/evidence=not experimental
complement(15683..16091)
/notes="match: GSS: Em:AQ031468"
complement(15788..16037)
/notes="match: STS: Em:G15194"
16015..16020
/genes="TCF21"
16041
/genes="TCF21"
17643..17736
/notes="WIR repeat: matches 49..156 of consensus"
repeat_region
18016..18073
/notes="29 copies 2 mer ct 72% conserved"
repeat_region
18787..18822
/notes="18 copies 2 mer ct 97% conserved"
repeat_region
19018..19113
/notes="MER5A repeat: matches 1..121 of consensus"
complement(19104..19320)
/notes="match: STS: Em:G21594"
complement(19114..19339)
/notes="match: STS: Em:G14783"
20283..20325
/notes="L2 repeat: matches 2675..2750 of consensus"
20892..21097
/notes="AluSq repeat: matches 1..207 of consensus"
21164..21277
/notes="MER21B repeat: matches 7..128 of consensus"
21279..21359
/notes="MER21B repeat: matches 274..361 of consensus"
21400..21807
complement(21885..22328)
/notes="match: GSS: Em:AQ416499"
complement(21975..22377)
/notes="match: GSS: Em:AQ005575"
22354..22941
22364..22813
/notes="match: GSS: Em:AQ486523"
match: STS: Em:G48886"
complement(22530..23061)
/notes="match: GSS: Em:AQ792775"
22898..23051
/notes="WIR repeat: matches 60..212 of consensus"
23399..23661
/notes="WIR repeat: matches 285..547 of consensus"
23662..23960
/notes="AluJb repeat: matches 1..301 of consensus"
23961..24243
/notes="WIR repeat: matches 1..286 of consensus"
24559..25036
complement(24819..25186)
/notes="match: STS: Em:G30455"
Query Match 56.4%; Score 717.2; DB 9; Length 25186;
Best Local Similarity 99.6%; Pred. No. 1.4e-135;
Matches 719; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 TCTACGGCCAGACTCTGGGAGTGGGAAACAGAGAGCCGGTCTCTCTGCTGCAGAGTC 60
Db 12915 TCTACGGCCAGACTCTGGGAGTGGGAAACAGAGAGCCGGTCTCTCTGCTGCAGAGTC 12974
QY 61 CTGCGGGTTCCTTCTCACAACCTCTCGAAGGGGAAAGGGTTGTGAGACCCACAGACCC 120
Db 12975 CTGCGGGTTCCTTCTCACAACCTCTCGAAGGGGAAAGGGTTGTGAGACCCACAGACCC 13034
QY 121 CAACTCCAGCTCCAGAGAGGTGGCTGCCCACTCGGAGAGCCCTCTTGGTTTCAGG 180
Db 13035 CAACTCCAGCTCCAGAGAGGTGGCTGCCCACTCGGAGAGCCCTCTTGGTTTCAGG 13094
QY 181 GTCTCTGTCTCTCTCACCCTCTTCTCGCTTTCTCTCTCTCTCTCTCTCTCTCTCT 240
Db 13095 GTCTCTGTCTCTCTCTCACCCTCTTCTCGCTTTCTCTCTCTCTCTCTCTCTCTCT 13154
QY 241 CTCTCTCCTCGTCCACTCCCCCAAAATGTCACCGGCTCCCTCAGGATGTGGAGAC 300
Db 13155 CTCTCTCCTCGTCCACTCCCCCAAAATGTCACCGGCTCCCTCAGGATGTGGAGAC 13214
QY 301 CTTCAAGAGGTGGAGATGTTGGAATGTGACGGGTTGAAATGGATTGGAACAAGGAATTT 360
Db 13215 CTTCAAGAGGTGGAGATGTTGGAATGTGACGGGTTGAAATGGATTGGAACAAGGAATTT 13274
QY 361 GTGACTTCCACGAGAGACCCAGAGAGCTCCAACTGCGAGAATGGGTCTCCCCAGAG 420
Db 13275 GTGACTTCCACGAGAGACCCAGAGAGCTCCAACTGCGAGAATGGGTCTCCCCAGAG 13334
QY 421 GGCCGCGGGCTGGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480
Db 13335 GGCCGCGGGCTGGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 13394
```







**This Page Blank (uspto)**